

Educational assessment for economies, societies and citizens

Towards a general theory of educational assessment

A thesis submitted to

The College of Education

Victoria University

in fulfilment of the requirements for the degree of

Doctor of Philosophy

by

Martinus Antonius Joseph Koomen

(Marten Koomen)

February 2019

Abstract

This study explores progress narratives in technological, social and personal dimensions and the effect different attitudes towards progress have on educational assessment design and use. By addressing this question this study works towards a general theory for informing design and use of educational assessment so that it better meets the needs of economies, societies and citizens. This study responds to rapid developments in technology that are affecting contemporary educational assessment design and use at a time when theoretical approaches are unable to frame these developments. It works towards identifying and exploring tensions emerging from different attitudes to progress, and how these attitudes inform contemporary educational assessment design.

This study engages with progress narratives through reconstructions of *historical materialism* associated with socialism (Marx & Engels, 1846/2000) and *creative-destruction* associated with capitalism (Schumpeter, 1942/2008). These formulations of progress narratives are explored through the work of Habermas (1981/1985, 1981/1992) and his use immanent critique (Antonio, 1981; Stahl, 2013). This sociological framing allows the effects of technological progress on educational assessment to be explored through its relationship with economies, societies, and citizens.

This study uses immanent critique and rational reconstruction to reconceptualise educational assessment for a contemporary context by coherently linking theories from sociology and educational assessment. Emphasis is given to educational assessment as symbolic media, through which several legitimation tensions are identified that are not currently addressed in educational assessment validation. These legitimation tensions have implications for educational assessment design as well as for education system management more generally.

Declaration of authenticity

I, Martinus Antonius Joseph Koomen, declare that the PhD thesis entitled *Educational assessment for economies, societies and citizens: Towards a general theory of educational assessment* is no more than 100,000 words in length including quotes and exclusive of tables, figures, appendices, bibliography, references and footnotes. This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work.

Signature _____ Date _____

Acknowledgements

I would like to acknowledge the support of my supervisors: Dr John Martino for providing principal supervision that at once gave welcome academic freedom while providing ongoing feedback that maintained a strong methodological focus in the critical theorist tradition, Dr Kirsten Sadler for providing timely insightful feedback to ensure the broader educational relevance of the work, and Dr Nathan Zoanetti for providing ongoing technical feedback on matters relating to contemporary educational assessment.

I would also like to acknowledge the support from Victoria University throughout my candidature, particularly Dr Marg Malloch and Associate Professor Dianne Hall for ensuring this research endeavour was kept on track. I would also like to acknowledge Professor Ron Adams and Associate Professor Deborah Zion for their insights during the introductory units of my candidature. I am also grateful for the scholarship awarded by Victoria University that allowed me to immerse myself in the study question and literature more than I might otherwise have been able to do.

I would also like to thank Ingrid Voorendt for proofreading an early draft of this dissertation and thank Danae De Kretser for proofreading the final draft. Both assisted in identifying textual errors before submission.

Table of Contents

Abstract.....	ii
Declaration of authenticity	iii
Acknowledgements	iv
Table of Contents	v
List of Figures.....	vii
List of Abbreviations.....	viii
Candidate Preface	1
Chapter 1. Introduction	9
1.1. The challenge of progress narratives	9
1.2. A pragmatic approach in the tradition of Habermas.....	14
1.3. Immanent critique and the anchoring polity of Victoria, Australia.....	22
1.4. This study.....	27
Chapter 2. Literature review	34
2.1. The affordances of technology.....	34
2.2. The malaise in educational assessment, contemporary critiques.....	54
2.3. The academic imperative for a pragmatic approach.....	60
Chapter 3. The methodology.....	65
3.1. Immanent critique and rational reconstruction	65
Chapter 4. Purposes of education and educational assessment.....	75
4.1. Pattern of three worlds.....	75
4.2. Purposes for educational assessment	81
4.3. Educational assessment and the pattern of the three worlds.....	89
Chapter 5. Educational assessment for systems and lifeworlds.....	100
5.1. From historical materialism to system and lifeworld	101
5.2. Educational assessment and symbolic media	108

5.3.	Colonisation and separation of system and lifeworld.....	112
5.4.	The active participant in the lifeworld.....	115
5.5.	Discourse ethics	122
5.6.	Deliberative democracy	129
5.7.	Rejoinder to postmodernism.....	133
Chapter 6.	Educational assessment for economies and polities.....	141
6.1.	The six markets of the AGIL scheme	141
6.2.	The integrative system – pragmatism and neoliberalism.....	147
6.3.	Markets as strategic action.....	156
Chapter 7.	Colonisation of educational assessment by science	168
7.1.	Educational assessment and psychological assessment.....	168
7.2.	The realist stance and ethics in educational assessment	179
7.3.	Performativity and the performative attitude.....	190
7.4.	Educational assessment for system management	200
Chapter 8.	Towards a reconstruction of educational assessment validity.....	210
8.1.	Normative validity	210
8.2.	The PISA and normative validity	223
8.3.	Legitimation tensions emerging from progress narratives.....	231
Chapter 9.	Discussion.....	243
9.1.	How ought educational assessment be designed	243
9.2.	Implications for educational assessment.....	252
9.3.	Looking ahead from the motivating issues	258
Chapter 10.	Conclusion	271
References	288

List of Figures

Figure 1 – Planning and review framework for assessment.....	37
Figure 2 – Three tiers of educational assessment.....	68
Figure 3 – Levels of use for educational assessment.....	82
Figure 4 – Perpendicular foci of educational assessment.....	92
Figure 5 – How schools relate to motivation and legitimation crisis tendencies	106
Figure 6 – Framework of social action used in this study	120
Figure 7 – The Toulmin approach to arguments	123
Figure 8 – The AGIL system and six markets.....	143
Figure 9 – The AGIL system with school differentiated from household.....	146
Figure 10 – Educational assessment and the integrative system.....	155
Figure 11 – The influence of public values, and private and commercial values	162
Figure 12 – Strategic management in school systems.....	203
Figure 13 – Legitimation tensions in educational assessment.....	232

List of Abbreviations

ACARA	Australian Curriculum, Assessment and Reporting Authority
ACER	Australian Council for Educational Research
AEA	Aptitude for Engineering Assessment
AERA	American Educational Research Association
AfL	Assessment for Learning
AGIL	Adaptive, Goal Orientation, Integrative, Latent Pattern Maintenance
ALSET	Australian Law Schools Entrance Test
APA	American Psychological Association
AQF	Australian Qualification Framework
ARG	Assessment Reform Group
AT21C	Assessment and Teaching of 21st Century Skills
ATAR	Australian Tertiary Admissions Rank
CSSE	Commonwealth Secondary Scholarship Examinations (Australia)
DNA	Deoxyribonucleic acid
ECD	Evidence Centred Design
IQ	Intelligence quotient
KSA	Knowledge Skill and Abilities
LAP	Learning Assessment Program (Victoria, Australia)
MBA	Master of Business Administration
NAPLAN	National Assessment Program – Literacy and Numeracy (Australia)
NCME	National Council on Measurement in Education (USA)
OECD	Organisation for Economic Co-operation and Development
PISA	Programme for International Student Assessment
SAT	A redundant acronym for a college admission test (United States)
TIMSS	Trends in International Mathematics and Science Study
UMAT	Undergraduate Medical Admissions Test
VCAA	Victorian Curriculum and Assessment Authority
VCE	Victorian Certificate of Education

Candidate Preface

The dissertation fulfilling this PhD project is a late career endeavour that seeks to contextualise a career rich in experience and professional development yet somewhat unsatisfied. The study question emerges from my professional experiences with debate over educational assessment spanning my career in education. While I have sought to provide an objective academic account, the material can evoke a sense of unease towards which this preface is directed.

Victorian audiences when presented with material from this dissertation tend to be at once welcoming and wary. The wary reactions centre on developments subsequent to the election of the Kennett Government in Victoria in 1992. This study only provides the minimum necessary detail of this period to enable the study question to be contextualised. However, the Kennett Government's mass closure of schools, shedding of teachers, and associated educational policy initiatives, profoundly affected the lives of many Victorian educators. For some, this study opens old wounds and triggers anxiety when personal narratives, or the established institutionalised ones, are not represented in the material. This can make the study hard to locate leading to questions over motivation and values, as well as questions over the authority I bring to speak on the matters addressed. It is towards these concerns that pertinent aspects of my career are described here.

As is generally the case, most teachers are influenced by their own experiences in education, and socialisation became the central focus of my approach. While I personally never suffered a lack of security or comfort, socialisation during my own schooling was adversely affected by rural isolation, a disrupted migrant background, and a troubled school environment. In response, socialisation became a central theme in my own work through focusing on activities that broadened student horizons to foster their connection to society.

I commenced teaching mathematics and physics in 1987 in an education system well-resourced and meaningfully engaged with complex ideas at both a school and system level. Students in the schools where I taught were mainly from socio-economically disadvantaged and migrant backgrounds. I struggled early in my teaching career but the senior teachers I encountered were competent and supportive. There was also support for the disadvantaged youth we taught with teachers having a distinct sense of agency that

entailed responsibility to advocate on behalf of students with fragmented lives and those from families involved in crime. School staffing levels were determined through a formula agreed to by teacher unions and government, and school staffing was locally administered through formally constituted teacher-led administrative committees.

Most of my teaching career was at a government girls school which was a contested space between conservative teachers and progressive feminists. I somewhat aligned with the progressives and was asked by the music teacher to develop a sound engineering and electronics program to support her own well-regarded performance music program. To establish my place in the school I embarked on a graduate diploma in technology education to retrain as a technology teacher. My teaching identity was formed teaching electronics to girls.

Some anecdotes about teaching electronics at a girls' school illustrate my ethos and attitude towards education. One such anecdote arises from a program we developed to enable girls to participate in a course delivered by a polytechnic that was a 20-minute train ride from the school. One girl from a migrant background and with a persistent behaviour problem expressed a strong desire to participate but was nevertheless apprehensive. She asked me to accompany her on the train to her first class at the polytechnic, which I did. On the train, I noticed the student was becoming anxious and apprehensive, I had never seen her so subdued. When we arrived at the class, she asked me to stay outside as she remained apprehensive, so I did. After about 10 minutes I looked in, then she came out and had returned to her usual boisterous self and told me in impolite terms that I was intruding and to go away. She felt independently comfortable in an electronics class with boys. It was among the most rewarding days of my teaching career, I knew that I had participated in a profound transition in a student's life, yet my role in that life had vanished. She did return to visit the school some years later after she had entered a career in electronics. She said I was nice enough, but a hopeless electronics teacher and she complained about the school's poor electronics facilities. She didn't care that only a handful of girls in the state participated in the program that led to her career in electronics, and that all those girls were from my class. This was a gratifying experience for me, I achieved something unique, and she had been empowered to demand what was equally her right.

During the 1990s there was an active interest across Victoria's education system in girls' participation in science and technology. Through my technology diploma I was inspired by the work of Judy Wajcman (1991, 2004) that was emerging around the theme of TechnoFeminism; not so much for political reasons but because it was fun. A particularly poignant story from my electronics class was that the girls who spent much time on their makeup, hair and nail polish, were also the best at the practical aspects of electronics. This was a surprise but consistent with the work of Wajcman that TechnoFeminism challenges how technology is used and considered. I enjoyed supporting girls exploring non-traditional contexts through excursions and through work-experience placements.

The feminists I encountered in my formative years talked about poststructuralism and how structures disadvantaged girls. As a technology and physics teacher I understood structures and focused on identifying and addressing structural barriers such as timetabling and logistics. I doubt I ever inspired any student or girl to do anything. I focused my work on structures which included exposing students to female role models. I considered inspiration a secondary concern to structures, and that teaching was more about providing students opportunity to emerge as citizens than to inspire them towards any personal ambition I might have for them.

My structural attitude is also illustrated when I was both school timetabler and middle school coordinator responsible for discipline. I was an unconvincing disciplinarian and at one time I had two particularly quarrelsome Year 9 classes. I found that I could minimise discipline problems for these classes through the structures of the timetable. I achieved this by ensuring that for period changeovers the two classes never crossed the school quadrangle thereby minimising opportunity for mischief. I also ensured that the next teacher for these two classes was always teaching close by so that the students would not be unattended for long after the bell rang. I remember enjoying going outside during period changeovers and listening to the school remaining quiet and calm. I also recall thinking that an educator's job is like that; when it is going well not much seems to be happening.

The election of the Kennett Government changed the environment dramatically in a way I didn't fully understand at the time. My feminist colleagues gradually left teaching with one becoming an author and another a goat farmer, while some are still close friends today

the changes fractured friendships and professional lives. State-wide teacher stop-work meetings that were once passionate and vitriolic somehow became calm and ordered. Something in the world had changed.

The changes initiated by the Kennett Government were stressful, and I embarked on a Master of Business Administration (MBA) to maintain sanity and to sustain my career. When discussing the MBA with friends at the time I told them it was a way to get to know the enemy. Further, the environment prior to the Kennett Government was also highly contested, and some of the changes the new government introduced made sense. There was also an emerging federal interest in the relationship between education, technology and knowledge-economies through the *Enterprising Nation* (Karpin, 1995) initiative that subsequently foundered when a new Conservative federal government was elected. Nevertheless, the MBA allowed me to get a better understanding of what was happening. I also completed a Master of Education through which I encountered approaches to organisational psychology, poststructuralist critiques, and Habermas. However, after completing both these qualifications my career in teaching was finished as my theory had outgrown the possibilities of practice.

Soon after leaving teaching I was selected for a junior role as research officer at the Australian Council for Educational Research working with the international management team implementing the Programme for International Student Assessment (PISA). The environment was familiar as I recognised key players involved in Victoria's school reform during the 1980s. I completed several university units on statistics and psychometrics on a part-time basis. I was tasked to look after the development of the PISA technical standards and to redraft them in response to advice from the project's technical advisory group, governing board, and technical experts. The American assessment standards were not suitable for providing guidance to PISA participants on matters such as sampling, verification of test materials, and standardising field operations. I was unofficially tasked with monitoring the quality of implementation across countries, so I was close to the process that excluded data from the Netherlands for PISA 2000, and data from the United Kingdom for PISA 2003. The nature of the arguments over data inclusion seemed familiar and reminiscent of debates over school staffing when I first started teaching.

As a junior research officer in the PISA, I was also responsible for managing two casual staff who were university students and who assisted with tasks such as collating feedback from countries on test material. One casual was a computer programmer, and on a day when we had nothing to do, I asked him if he could program a PISA test item on a computer and to add a video. Subsequent to this small act of bricolage, it seemed like no time before the first PISA computer-based assessment was in a field trial in twelve countries. The developers of the software, videos, and animations were either still university students or new graduates. I understood my role as manager as ensuring task clarity and maintaining clear communication across software developers, animators, video producers, test developers, field operation experts, and psychometricians. I managed the computer-based project and met expectations in terms of timeline, budget and operations. The quality of the data was good, but the scaling of the data was a problem. My instrumental systems thinking had run its course.

Politically, the PISA 2006 computer-based assessment of science project was out of my depth and I began to encounter issues of profound significance. First, through my work on the PISA technical standards I had a thorough understanding of the PISA testing procedures and the requirements for valid and reliable measures. The computer-based tests were administered on standardised laptop computers that national testing centres were required to bring to schools. This ensured uniform testing conditions across students, schools and countries. All students doing the computer-based test also did the paper-based test in standardised conditions. While this produced reliable comparable data, it also encountered political resistance from countries due to the cost and logistical effort supplying and transporting laptops. These political forces suggested to me that the PISA was not about educational research and furthering knowledge through scientific principles, but about something else.

Second, the scales obtained through the computer-based test were comparable with the paper-based test but incompatible for the purpose of making a single scale. For example, I led the development of an interactive stimulus that simulated a nuclear power-plant with two interactive controls. The PISA science experts particularly liked this stimulus as it assessed aspects of scientific literacy not able to be assessed through paper-based testing. However, this stimulus was included in items that demonstrated bias in favour of girls who found the item easier than boys. This bias was at variance with the paper-based test.

In Victoria during the 1980s, we understood that girls responded differently to assessment, and that technology changed the nature of assessment. The bias towards girls in the nuclear power-plant item made me wonder if my earlier interest in TechnoFeminism might have influenced development of the item's stimulus. Nevertheless, it became apparent to me that technological progress and the perspective of gender could not be accommodated by the PISA scales. The OECD published a separate report for this first excursion into computer-based testing.

A third issue was that my success implementing a technology-based project led to commercial pressures and pressures from other egos. The software team of students I had built had inadvertently become a competitor to the organisation's internal software department, causing tensions. I found myself in strange and unusual circumstances. I remember being invited to lunch by a Russian contingent on the river Danube in Bratislava, and at another time being invited to Menlo Park in Silicon Valley to talk with people from Stanford University. I retell this not by way of boasting, but out of embarrassment as I didn't know what I had to offer. I remember after my Silicon Valley visit seeking respite in a book shop in the Haight-Ashbury district of San Francisco to read about philosophy. While I didn't have the words to describe it, I found the position of the PISA towards computer-based assessment untenable. Again, friendships fractured, and I left the PISA project to be closer to the practices of teaching to take up a position as Human Resource (HR) manager for Victoria's Department of Education and Early Childhood Development (the Department).

HR was not considered a glamorous division within the Department, particularly as there was a high profile Ultranet project, a statewide technology platform with an assessment component, emerging from another division in the Department. Pertinent to this dissertation, it is through the function of HR that the teaching workforce as a matter of routine interacts with systems of law through selecting and terminating teachers, as well as interacting with economic systems through the payroll. I led a team of 40 staff to manage moving Victoria's government schools onto a new payroll system at the same time as the Ultranet project was being conducted. My most rewarding moment as a public servant was when I led three of my team to the teacher union headquarters to address a meeting of union delegates who had concerns over the new payroll system. It was a tense meeting as they were apprehensive, and I had been in that same room when I was a teacher

union member arguing against the government. However, we managed to answer all the delegates' questions to their satisfaction, and they gave us a clap and a cheer when we left. The new payroll system rolled out silently and with little fanfare or problem.

Internally the rollout was tense as payroll systems are complex. There was a good team working across the Department, but random issues arose from time-to-time. It was difficult to discern friend from foe, and only a commitment to the task by friends led to the payroll project succeeding. It was only later when the Ultranet project failed and became the subject of a public corruption inquiry that the distinction between friend and foe became clearer. Nevertheless, by this time friendships and professional lives had again been fractured.

I took some leave to develop my own online assessment system. I was successful and able to attract backing from angel funders, but their demands for returns made any project unworkable. Further, I understood that good test developers are at the heart of educational assessment and my potential backers did not have the capacity to support the endeavour in any meaningful sense other than financial. I returned to work in various policy roles in the Department, which included giving advice to other failed attempts at technology-based educational assessment projects. However, I found that I had run out of theory to inform progress, and that the world seemed to have run out too. This motivated me towards a PhD.

The material in this dissertation draws on my studies and experience in education, technology, management and psychometrics. It is written from the perspective of an aloof structuralist school timetabler, and not from the perspective of a caring empathetic year-level coordinator. I understand the tragedy for students associated with a decline on the PISA scales. I know for some students a decline in the PISA scales means it will be a little more difficult to read a train timetable, for others it will be a little more difficult to determine which medicine to administer, for others again it will be a little more difficult to fly a plane. These outcomes are an emotional tragedy, a tragedy to which this dissertation might appear to lack empathy. However, through a focus on structures I aspire to enable teachers to more meaningfully engage with students so that they emerge as engaged citizens.

The dissertation may also come across as unfamiliar as it does not sit in an immediately recognisable academic or familiar tradition. My approach to educational practice draws on Wajcman, Piaget, and Dewey among others. I also draw on postmodernists such as Derrida and de Certeau in my personal approach to life. When shopping, I sometimes maximise like Friedman or Hayek might. Therefore, some of the terms used in the dissertation – collectivism, capitalism, neoliberalism and postmodernism – are used for utility of argument and not for defining a personal position, these terms reflect how protagonists refer to each other. My work has always been pragmatic in an everyday sense. This study takes a formal pragmatic approach for an academic context. The Master of Education introduced me to the psychological concept of splitting which is associated with projection and identification that can lead to toxic dichotomies between opposing positions. Toxic dichotomies can be offset through meaningful engagement and this dissertation seeks to pursue its question through pragmatic meaningful engagement. Nevertheless, audiences are often more comfortable when a recognisable position clearly opposed to another is presented.

Finally, unease towards this dissertation might arise from its inherent design focus. Kress (2010) describes critique as exploring the present effects of past superiors. This dissertation embraces the design and forward-looking approach advocated by Kress (2010). For some, design means revolution, for others a curb on freedom, and for others again a path towards emancipation. With design comes the fear of change and loss, but also purpose.

While this study is motivated by personal experiences these are not directly drawn upon. Only passing references are made to the Ultranet project and to the PISA computer-based assessment of science, for example. This study is not an ethnographic or phenomenological account, and personal anecdotes do not provide warrants for this study's argument. This is a formal academic study in the pragmatic tradition that draws on the academic literature and, where appropriate, publicly available reports. Nevertheless, I can warrant that each concept described herein pertains to issues I have personally faced in my career in education. Each can be supported by a meaningful anecdote.

Marten Koomen

Chapter 1. Introduction

This chapter introduces this study's research question that addresses how different attitudes towards progress narratives affect educational assessment design and use. Progress narratives are framed throughout this study around themes of increasing affordances of technology, increasing demands on social norms for social cohesion, and increasing personal individuation in tastes and wants. The study question is addressed with the view to informing contemporary educational assessment design so that it better meets the needs of economies, societies and citizens. New possibilities afforded by technology provides the main imperative for this study, which is also motivated by issues arising in the context of Victoria and Australia.

The first section describes the context of the study question and sociological framings from which it emerges. Three issues are identified from the context of Australia that suggest contemporary issues towards progress narratives in educational assessment. The second section describes the pragmatic approach taken by this study and the theoretical framework adopted. The third section provides an overview of developments around educational assessment in the state of Victoria and Australia which provides the contextual basis for this study's pragmatic immanent critique.

The fourth and final section describes three key contributions this study seeks to make to the academic literature: broadening educational assessment validation to include legitimacy concerns; enhancing the conceptualisation of educational assessment as symbolic media involving two interpretations; and conceptualising educational assessment as symbolic media with a linguistic and a numerical component. Several corollaries emerge from these contributions which are also discussed before the section concludes by describing the structure of the following chapters.

1.1. The challenge of progress narratives

This study responds to rapid developments in technology that are affecting contemporary educational assessment design and use at a time when theoretical approaches are unable to frame these developments. It works towards identifying and exploring tensions emerging from different attitudes towards progress, and how these attitudes inform contemporary educational assessment design.

Koomen and Zoanetti (2018) address how technological progress can affect educational assessment design and use, this progress narrative is located within a broader academic tradition sometimes referred to as the dialectic in history (Antonio, 1981; Hegel, 1807/1977; Marx & Engels, 1846/2000; Plato, 1963). Progress narratives were challenged by Lyotard (1979/1984) through what Jameson (1984) described as an “incredulity toward metanarratives” (p. xxiv.). Similarly, Fukuyama (1992) announced “the end of history” to signal the end-point of humanity’s sociocultural and ideological evolution. An attitude against progress narratives is also reflected in an assumption that wants are fixed in contemporary economics (Friedman, 1962/2008, p. 13). This fixed view of society is further reflected in the linking and equating methodologies of educational assessment that focus on measurement invariance so that educational constructs remain the same over time (Dorans, Moses, & Eignor, 2011; Gebhardt & Adams, 2007; Holland & Dorans, 2006; M. von Davier et al., 2019). This study explores tensions between attitudes that embrace progress and attitudes eschew it.

This study responds to evidence that contemporary approaches to educational assessment are not meeting the needs of society and considers this evidence from the perspective of progress narratives. Three specific issues are identified: a sense of malaise in the field of educational assessment (Masters, 2013), issues with implementing technology-based educational projects (IBAC, 2017; Pearson, 2012), and Australia’s performance on the Programme for International Student Assessment (PISA) scales (Thomson, De Bortoli, & Buckley, 2013; Thomson, De Bortoli, & Underwood, 2017). This study works towards providing an academic framing for these issues with the aim of promoting better educational assessment design and use.

Sense of malaise in educational assessment

This study identifies a malaise in educational assessment and explores this malaise from the perspective of progress narratives. This malaise is expressed by Masters (2013) through the observation that

the field of educational assessment is currently divided and in disarray. Fault lines fragment the field into differing, and often competing philosophies, methods and approaches. At the same time, there are unprecedented external pressures for assessment reform. (p. 1)

Masters (2013) goes on to describe the field as divided by a range of default dichotomies, such as “quantitative versus qualitative; formative versus summative; norm-referenced versus criterion/standards-referenced” (p. 2) among others.

A further sense of malaise is reflected in a more recent high-profile report from Gonski et al. (2018) that identifies deficiencies in Australia’s educational practices where

constraints include inflexibility in curriculum delivery, reporting and assessment regimes, and tools focussed on periodic judgements of performance, rather than continuous diagnosis of a student’s learning needs and progress. (p. ix)

The report goes on to make recommendations for an “online and on demand student learning assessment tool for teachers for the purposes of formative assessment and tailored teaching” (p. xii). This recommendation can be considered as an imperative for educational assessment to better engage with the progress narrative of technology.

This study explores the sense of malaise in educational assessment through several themes including an inordinate focus on the numerical component of educational assessment (Ball, 2015; Biesta, 2015b; Grek, 2009; Lingard, 2011). This numerical focus in turn leads to heightened empirical use of educational assessment data for strategic and goal-oriented action and the rise of markets (Hogan, 2012, 2016a; Hogan, Sellar, & Lingard, 2015, 2016). The presence of goal-oriented and market-oriented actors might, for example, explain the unhelpful dichotomies observed by Masters (2013), as marketing techniques encourage commercial vendors to differentiate themselves from others (Kotler, 1972; Kotler & Levy, 1969). The effect of markets and goal-oriented activity in educational assessment design and use is further explored in Chapter 6.

How educational assessment is used across society provides another possible explanation for the malaise described by Masters (2013). Masters (2013) argues in favour of a unifying principle

that *the fundamental purpose of assessment is to establish where learners are in their learning at the time of assessment.* (pp. 5-6)

This proposed principle suggests that once the location of a learner is identified on a line, plane, or multidimensional space, details of that location can be used to inform activity

across society without further argument. This position contrasts, for example, with the position of Newton (2012, p. 267) who argues that it is wrong for a test result to be given universal licence for any purpose, and that each use of a test result requires its own validity argument. Master's (2013) principle frames educational progress from a perspective of numerical achievement levels, while ignoring that test content from which measures derive might be evolving. This study explores whether it is appropriate to simply consider the numerical component of educational assessment data, or whether numerical data needs to be considered within broader narratives of progress in society that are expressed pragmatically through language.

Problems implementing large-scale educational technology

Difficulties that education systems have implementing large-scale technology-based projects further suggests issues with progress narratives in educational assessment. One example of a poorly implemented technology-based educational project is the Ultranet project in the Australian state of Victoria (P. Griffin & Woods, 2006; IBAC, 2017; Pearson, 2012). The Ultranet project was designed to deliver a Victoria-wide technology platform for activities including curriculum planning, curriculum delivery, educational assessment, reporting, and attendance monitoring (P. Griffin & Woods, 2006). Victoria's auditor general found that the "Ultranet was poorly planned and implemented" (Pearson, 2012, p. 19). A subsequent corruption inquiry found that the Ultranet project was closed down in 2013, seven years after it had commenced, and "was likely to have cost somewhere between \$127 million and \$240 million" (IBAC, 2017, p. 7). The Ultranet sought to modernise Victoria's education delivery through technology but instead wasted significant funds for no educational benefit.

The Ultranet episode is not an isolated example with Australia's National Assessment Program – Literacy and Numeracy (NAPLAN) also experiencing issues, albeit less dramatic. The Australian Government allocated significant funding in 2014 to enable the NAPLAN test to be available online in 2016 (Pyne, 2014). Yet in 2017, state governments across Australia announced that they would not be participating in the online delivery due to technical issues with the platform (Ellery, 2017; Merlino, 2017; Ramsay, 2017). NAPLAN, and particularly a limited rollout of NAPLAN online, generated further concerns in 2018 over the comparability between paper-based versions of the NAPLAN

test and online versions (Merlino, 2018; Perelman, 2018; Robinson, 2018). The transition to NAPLAN online also points to issues of progress narratives and educational assessment.

Australia's decline in educational achievement

Australia's reported decline in educational achievement over successive cycles of the PISA (Thomson et al., 2013; Thomson, De Bortoli, & Underwood, 2016) points to another curious issue with progress narratives. Australia's decline is of particular interest given that Australians led the development of the PISA, developments further described in the third section of this chapter (Adams & Wu, 2002; McCurry, 2017; OECD, 2000).

One explanation for Australia's decline explored through this study is a shifting focus in system coordination from a pragmatic language-based approach to a numerical abstract approach. This study observes a shift to numerical coordination occurring over several decades. A pragmatic focus is evident, for example, in an account given by McCurry (2017), who describes Australia's capacity in educational assessment item-writing emerging from the 1960s. Further, Barcan (1996, 2003) describes extensive development in school-based curriculum from the 1970s through to the 1990s across Australia. Both the accounts of McCurry (2017) and Barcan (1996, 2003) suggest a significant emphasis on pragmatic coordination of educational activity from the 1960s through to the 1990s. This study observes a shift towards numerical coordination as Australia's capability in numerical methods emerged around the 1980s (Adams, Wilson, & Wang, 1997; Andrich, 1978; Masters, 1982). Contemporary critique suggests an inordinate emphasis on numerical data in the formulation of educational policy (Biesta, 2015b; Gillis, Polesel, & Wu, 2016; Gorur & Wu, 2015; Grek, 2009; Lingard, 2011). This study explores whether Australia's decline on the PISA scales can be explained by a disengagement with progress narratives caused by a shift in focus from pragmatic to numerical coordination in education systems.

1.2. A pragmatic approach in the tradition of Habermas

This study takes a pragmatic approach towards its research question through the tradition of Habermas (1998). Habermas (1998) describes the task of universal pragmatics as identifying and reconstructing "universal conditions of possible mutual understanding" (p. 21). Huang (2014) considers the pragmatic tradition notoriously difficult to define and

only makes a passing reference to Habermas when describing developments within the tradition. Huang (2014, pp. 1-3) approaches pragmatics as the systematic study of meaning through language, with Habermas' (1998) approach used as the framework for this study.

Bacon (2012) associates Habermas with pragmatism characterised as a matter-of-fact approach to problem solving that addresses tangible and practical aspects of life. This practical form of pragmatism is important in educational assessment as it is an inherently a practical activity subject to resource and timeline constraints. This study addresses its question through pragmatics as described by Huang (2014) and pragmatism as described by Bacon (2012) through the tradition of Habermas (1998).

In adopting the pragmatic approach of Habermas (1976/1979, 1981/1985, 1981/1992, 1992/1998), this study is able to draw on a tradition that traces back to the eighteenth-century through thinkers such as Kant, Hegel and Marx (Bowie, 2003; de Berg & Large, 2012). This tradition focuses societal coordination, universal principles, and intersubjectivity. It is a tradition that engages with progress narratives, including *historical materialism* developed by Marx and Engels (1846/2000) that addresses how technology changes the modes of production in society. The tradition includes Hegel's (1807/1977, 1816/1998) dialectical method of argument that emerges from engaging with different logical conceptions of consciousness and progress, a tradition that builds on the dialectical method of Plato (1963) in the *Republic*. The dialectical tradition of engaging opposing and contradictory logics is drawn upon throughout this study.

By focusing on progress narratives this study takes a dialectical approach that is distinct from traditional approaches. Schumpeter (1942/2008, pp. 167-168), for example, associates the terms *collectivism*, *socialism*, *centralism* and *communism* with Habermas' tradition to build an economic argument framed through the "traditional contrast between capitalism and socialism" (p. 167). This study does not directly engage with this traditional capitalist and socialist contrast, and instead engages with attitudes that consider progress in society as either fixed or dynamic. The work of Schumpeter (1942/2008) and Marx (2000), for example, are traditionally considered as opposing perspectives through a capitalism-socialist divide however both engage with progress narratives (J. E. Elliott, 1980; Foster, 1983; Rahim, 2009). Instead of engaging a dialectic

between capitalism and socialism this study engages the dialectic between those taking a fixed view of society and those engaging with a dynamic society through progress narratives.

The pragmatic approach of Habermas (1976/1979, 1998) engages with progressive narratives and contrasts to both neoliberal and postmodern traditions that turn away from technological and social progress in their own ways. Habermas (2001) observes that for

different reasons, postmodernism and neoliberalism ... ultimately share the vision of the lifeworlds of individuals and small groups scattering, like discrete monads, across global, functionally coordinated networks, rather than overlapping in the course of social integration, in larger, multidimensional political entities. (p. 88)

It is the distinction that Habermas (2001) makes between his tradition and that of postmodernism and neoliberalism that frames the pragmatic approach for this study.

Habermas' (1976/1979) pragmatic approach builds on the tradition of Marx (2000) and might also be described as socialist. There are further alternative formulations with Habermas (1997) associating himself with the modernist and enlightenment traditions. The tradition is also associated with structuralism (Piaget, 1968/1973). While this study builds on all these traditions, it maintains a focus on educational assessment and does not engage with broader economic dynamics attributed to capitalism by the socialist tradition, such as issues of wealth distribution and class structures explored by Marx (2000). Instead, this study addresses a small subset of concerns focusing on educational assessment design and use through the pragmatic tradition of Habermas (1998).

Progress narratives provide the point of departure for this study and these narratives are described in various ways (Nisbet, 1988/2012, 1969/2017). Durkheim (1933/2012), for example, uses the concept of *division of labour* to explore changes to modes of production (Smith, 1776/1993). The theme of division of labour is explored in the contemporary context of evolving digital technology and its effect on work by Levy and Murnane (2004). The rise of digital technology and its effect on social, cultural and economic developments, is also explored by Drucker (1968/1992) through discontinuities that emerge in *knowledge-based* economies, with knowledge considered in this study as the

basis for contemporary modes of production. The effect of technological progress on the consciousness of children and on learning is explored by Vygotsky (1978). Papert (1993) also explores the impact of digital technology and computers on educational practices and the learning patterns of students. This theme has more recently been explored by Prensky (2001) through the notion of *digital natives*, a conceptualisation challenged by Kirschner and De Bruyckere (2017). This study does not so much seek to describe or conceptualise narratives of progress other than those related to the implementation of educational assessment. Instead, this study explores orientations and attitudes towards progress and how tensions between attitudes affect educational assessment design and use. Pragmatism is characterised as addressing progress through ethical reasoning, neoliberalism through markets, and postmodernism as turning away from progress.

Pragmatics

This study's pragmatic approach reveals a dialectic between a linguistic component of educational assessment that addresses meaning and a numerical component addressing educational measurement (Brennan, 2006b; Messick, 1989; Thorndike, 1971). This dialectic is reflected in the approach of Wu (2014) who argues that statistics alone cannot provide hard evidence, and that interpretation of educational assessment data relies on personal experience and sense-making. Habermas (1981/1985, 1981/1992) locates sense-making in a *lifeworld* which is elaborated by Husserl (1954/1970, p. 106) as a world of sensing, intuition and appearances. As lifeworld horizons increase and become increasingly complex, Habermas (1981/1992) draws on Parsons (1963a, 1963b) to elaborate symbolic media as way of communicating meaning. A central concern that emerges from the critique of Habermas (1981/1992) is the maintenance of a meaningful connection between the linguistic and numerical components of educational assessment as symbolic media.

Habermas (1976/1979) considers society as progressing along technological and social dimensions, and his work addresses tensions that progress generates for symbolic media such as educational assessment. These tensions are explored by Habermas (1983/1996, 1992/1998) through *discourse ethics* and *deliberative democracy* that respectively focus on pragmatic interpersonal communication and how it is institutionalised.

Habermas' (1998) pragmatic approach is distinguished from postmodernism by seeking to address social progress through universal principles of moral and ethical reasoning. Habermas (1996/2000), in the *Inclusion of the Other* for example, considers that increased moral development provides the capacity to address issues of identity. Challenges to Habermas' (1998) tradition particularly emerge around identities of gender (Butler, 1988; Irigaray, 1985; Wittig, 1993), race (Gandhi, 1998; Said, 1978/1994; Spivak, 1985/2010), and sexuality (Foucault, 1984/1988, 1976/1990, 1984/1990). In seeking to address these challenges through a universal synthesis of cognitive-instrumental and moral-practical reasoning (Habermas, 1996/2000), pragmatism is distinguished from postmodernism as described later in this section.

Neoliberalism

Neoliberalism, like pragmatism, seeks to address progress emerging from technological and social development in society. However, where Habermas' (1998) pragmatic approach proposes deliberative democracy through institutions to address progress (Habermas, 1992/1998), neoliberalism eschews institutions for central planning and its associated bureaucracies (Hayek, 1944/2007, 1935/1963; von Mises, 1944/2007). Instead, neoliberalism addresses societal coordination through markets "led by an invisible hand to promote an end which was no part of his (sic) intention" (Smith, 1776/1993, p. 292). Neoliberalism's preference for markets over government is its central distinction from Habermas' (1998) approach (Friedman, 1962/2002).

This study distinguishes between neoliberal philosophies and its associated technical monetarist policies to focus on the latter. Stedman Jones (2012) observes that "Friedman himself insisted that his technical and empirical work as an economist was distinct from his political philosophy and activism" (p. 5). Here, Friedman (1953/2008, 1969/2008, 1962/2008) is insisting that his approach to monetarist economics is distinct from philosophies towards education and other aspects of human life such as those expressed in *Capitalism and Freedom* (Friedman, 1962/2002). Friedman (1962/2002), for example, advocates for a market-based approach in education. Neoliberalism as an ideology is not critically addressed in this study, other than in ways that this study identifies neoliberalism and monetarist policy directly affecting educational assessment design and use.

This study does not seek to make a totalising judgement on the efficacy of neoliberalism and other ideologies in coordinating society more broadly. For example, Stedman Jones (2012) describes neoliberal policy as increasingly being adopted as a result of a series of 1970s oil and financial crises that were unable to be solved through the then Keynesian (1953/1964) economic approaches. While Stedman Jones (2012) identifies neoliberal thought emerging from the 1940s and propagated by coteries such as the Mont Pelerin Society (Stedman Jones, 2012, p. 31), there is also a sense that the adoption of neoliberal policy resulted from economic and political imperatives. Stedman Jones' (2012) account suggests that these imperatives arose out of complexity emerging from progress in society, particularly during the 1960s. These developments suggest that neoliberalism and its associated monetarist economics may have some efficacy in coordinating certain spheres of society.

This study further observes that neoliberalism began to affect Australian education around 1992 when the Kennett Government was elected in Victoria (Laming, 2012; Marginson, 1993, 1997a; Spaul, 1999). This study also notes that the Organisation for Economic Cooperation and Development (OECD, 2017a) reports “an impressive twenty-five consecutive years of output growth” (p. 16) for Australia since that time. Therefore, while this study challenges neoliberal approaches towards educational assessment, it does not seek to make a totalising judgement on neoliberalism and its associated monetarist economics.

Of interest to this study are the assumptions that underpin neoliberal policy, or more explicitly, monetarist economics or positive economics (Friedman, 1953/2008), and its associated empirically-based numerical techniques and models. The empirically-based monetarist models of Friedman (1969/2008, p. 48) are based on four types of capital, of which one is *human productive capital*. This form of capital is empirically modelled by G. S. Becker (1964/1993) through *human capital theory*, a theory motivated by a desire to “estimate the money rate of return to college and high-school education” (p. 29). Human capital theory links investment in education with increased earning capacity of citizens, and links investment in education with the economic growth of countries (G. S. Becker, 1964/1993; Hanushek, 2013). Marginson (1993) identifies human capital theory as a driver of educational policy among member countries of the OECD since the 1960s. Educational assessment is important to human capital theory as it provides an empirical

link between investment in education, the outputs of education, and economic returns. The empirical link between human capital development and economic growth are explored, for example, by Hanushek (2013) using educational assessment data. This study explores the monetarist assumption that economic growth and its relationship to education refers to an unchanging society where technological and social progress is ignored in its models.

This study's particular interest in monetarist economics focuses on an assumption made by Friedman (1962/2008) on how wants relate to technological and social progress. Friedman (1962/2008) assumes that

economic theory proceeds largely to take wants as fixed. This is primarily a case of division of labor. The economist has little to say about the formation of wants; this is the province of the psychologist. (p. 13)

Pollak and Wales (1992, p. 124) argue that Friedman (1962/2008) is here expressing a dominant view among economists, and this study explores how this assumption might influence educational assessment design and use. The notion that wants are considered fixed is interpreted here as a disengagement with aspects of progress narratives. This study explores Friedman's (1962/2008) assumption and it affects the use of educational assessment data in economic modelling and planning. The wants as fixed assumption in economics is associated in this study with the constructs in educational assessment being similarly fixed for trend reporting (Dorans et al., 2011; Gebhardt & Adams, 2007; Holland & Dorans, 2006; M. von Davier et al., 2019). The relationship between wants as fixed and educational constructs as fixed makes Friedman's (1962/2008) economic assumptions of interest in the context of this study's question.

Poststructuralism and postmodernism

An imperative for this study to engage with the postmodernist tradition emerges from a bifurcation in Habermas' (1981, 1997) tradition that occurred in the later part of the twentieth century (Rorty, 1984, 1995). It is through this bifurcation that Lyotard (1979/1984) developed the concept of performativity in a distinct tradition to Habermas (1981, 1997). The concept of performativity was appropriated by Ball (2003) and drives much contemporary critique of educational assessment (Hardy & Lewis, 2016; Keddie, 2016; Lambert, Wright, Currie, & Pascoe, 2015; Solomon & Lewin, 2016). Habermas

(1981, 1997) distances himself from thinkers such as Lyotard (1979/1984), as well as from thinkers such as Foucault (1966/2002) and Derrida (1966/2007). Where Habermas (1997) considers modernity an unfinished project, others eschew the aims of modernity and advocate for postmodernism which Jameson (1984) associates with an incredulity toward progressive narratives. This bifurcation is sometimes cast in terms of structuralism (Piaget, 1968/1973) versus poststructuralism (Butler, 1990/2007; Harcourt, 2007), or in terms of a *postmodern condition* (Lyotard, 1979/1984), or a *condition of postmodernity* (Harvey, 1990). This study engages with this bifurcation through the perspective of pragmatism opposed to postmodernism.

While this study works in and defends Habermas' (1998) pragmatics and associated structuralist (Piaget, 1968/1973) approaches, poststructuralist critiques nevertheless pose justified challenges. Wittig (1993), for example, challenges Marxism (2000) on the basis that it does not address the oppression of women. Foucault (1984/1988, 1976/1990, 1984/1990) challenges traditional attitudes towards sexuality. Derrida (1966/2007), Spivak (1985/2010) and Said (1978/1994) charge structuralism as being centred on European culture. Crenshaw (1991), in a similar fashion, addresses the oppression of black women in the United States through the concept of *intersectionality*. This study considers these challenges to Habermas' (1998) tradition valid and engages with issues of gender, sexuality and race throughout as appropriate.

While this study acknowledges the validity of poststructuralist critiques of modernism and structuralism (Butler, 1990/2007; Derrida, 1966/2007), it differentiates between pragmatic and postmodern solutions. For example, where Habermas (1983/1996) argues in favour of universal pragmatic consensus, Lyotard (1979/1984), in the postmodern tradition, considers consensus “an outmoded and suspect value” (p. 66). To differentiate critiques of modernism and structuralism from postmodern solutions, this study uses the terms poststructuralism to refer to critiques of modernism and structuralism, and postmodernism to refer to solutions that contrast to Habermas' (1981) pragmatics.

Postmodernism

Habermas (2001) considers postmodern critiques conservative and presaged on modernism, “as if the material structure of society were made up of the concepts and discourses of social scientists” (p. 147). Where, “[o]n the basis of modernistic attitudes,

they justify an irreconcilable anti-modernism” (Habermas, 1981, p. 13). This study finds postmodern critique of educational assessment similarly conservative, with postmodern critiques generally based on modernist concepts in a way that fails to address the dialectical tension between those concepts and practices in society, which results in a failure to engage with progress narratives.

The critique by Ball (2003, 2013) of educational assessment, which draws on Lyotard (1979/1984) and Foucault (1975/1991, 1978/1991, 1966/2002), provides one example of a conservative postmodern critique. Ball (2003) provides the “three interrelated *policy technologies*; the market, managerialism and performativity” (p. 216) as a central focus of his critique. However, Ball (2003) does not engage with these technologies in terms of methods and theories, and how these apply to education. Ball (2003) does not engage with Lyotard (1979/1984) or Austin (1962/1975) as the seminal developers of the concept of performativity, and instead claims the concept as his own

What do I mean by performativity? Performativity is a technology, a culture and a mode of regulation that employs judgements, comparisons and displays as means of incentive, control, attrition and change based on rewards and sanctions. (Ball, 2003, p. 216)

In defining performativity in his own terms, Ball (2003) forgoes the possibility of talking back and being in pragmatic dialogue with the concept of performativity as it evolves alongside other progress narratives in society. This makes the critique of Ball (2003) conservative as it does not seek to develop the concepts that it engages with. Like Ball (2003), this study considers performativity important for educational assessment and substantively addresses the concept in Chapter 7 by drawing on Austin (1962/1975), Lyotard (1979/1984), and Butler (1988, 1990/2007).

One further example of postmodern critique is the extensive critique Ball (2016a, 2016b) provides of neoliberalism. Here too, Ball (2016a, 2016b) does not engage with neoliberalism’s seminal thinkers such as Popper (1945/2002), Hayek (1944/2007), von Mises (1944/2007) or Friedman (1962/2002) (Stedman Jones, 2012). Ball’s (2016a, 2016b) critique of neoliberalism is postmodern as it is presaged on a vague interpretation of the concept of neoliberalism without seeking to engage or develop the concept. This

study differentiates itself from Ball's (2016a, 2016b) critique by engaging with seminal neoliberal thinkers.

This study's methodology of immanent critique and rational reconstruction is described next.

1.3. Immanent critique and the anchoring polity of Victoria, Australia

Habermas' (1998) pragmatic tradition is associated with three broad methodologies: immanent critique (Stahl, 2013), rational reconstruction (Carnap, 1928/2005), and critical theory (Antonio, 1981; Kellner, 1992, 2003). This study concurrently employs the methods of immanent critique and rational reconstruction which are further elaborated in Chapter 3. Both immanent critique and rational reconstruction address practices immanent to communities and societies. Habermas (1976/1979) accesses immanent practices for rational reconstruction through the notion of the *competent subject*. Immanent practices can also be accessed by investigating societal practices and interpretation of those practices by theorists (Stahl, 2013). While the way immanent practices are addressed may vary, immanent critique is historically and contextually located and draws on contemporary practices.

This study is contextually based in the state of Victoria in Australia, a jurisdiction which this study uses to identify contemporary practices and approaches. Victoria is selected because it is the context in which this study is undertaken and is the context for this study's motivating issues. Victoria also provides a good example as the state has been central to key developments in educational assessment, however this study does not seek to ignore parallel and preceding developments in other jurisdictions. Instead, Victoria, and Australia, are used as vehicles of convenience for providing a narrative for developments occurring more broadly around the globe.

Educational assessment in Victoria, Australia

Australia provides a fruitful anchoring context for a study on educational assessment as it has, at times, been at the leading edge of developments. McCurry (2017), for example, describes the evolution of Australia's item-writing capability as emerging out of Australia's *Commonwealth Secondary Scholarship Examinations* (CSSE) that commenced in 1964. McCurry (2017, p. 124) argues that the style of items found in the

CSSE were a progenitor of the PISA items. This is a justifiable claim, not only because both the CSSE and the PISA were established through the same organisation, the Australian Council for Educational Research (ACER) based in Victoria, but also because of the distinctive cross-curricular stimulus-based nature of both tests. The stimulus-based items that McCurry (2017) illustrates from the CSSE share a similar form to those found in the PISA tests (OECD, 2000). Australians have also led in other areas of educational assessment.

Australians have been prominent in developing theory and software related to *item response theory*; particularly developments of the Rasch (1960/1980) model. The Australian Geoff Masters (1982) developed a *partial credit model* based on the work of Rasch (1960/1980). Masters was also the inaugural chair of the PISA *Technical Advisory Group* (Adams & Wu, 2002). Furthermore, Adams et al. (1997) developed the *mixed-coefficients multinomial logit model* as a generalised development of the work of Rasch (1960/1980). This model was implemented by Wu, Adams, and Wilson (1997) in the *Conquest* software that was used in the PISA (Adams & Wu, 2002). These innovations were all used in the establishment of the PISA for which the inaugural project directors were Adams and Wu (Adams & Wu, 2002). Australia's NAPLAN, which is administered by the Australian Curriculum, Assessment and Reporting Authority (ACARA), also employs methods developed by Australians and used in the PISA (ACARA, 2014, p. 2). In addition to Australians leading in test development and item response theory, others have led in the development of assessment content.

Australians have influenced global developments in curriculum, or more specifically, educational assessment domains and frameworks. Curriculum leadership emerged in Australia during the 1980s, a period Barcan (2003) characterises as a time of significant national curriculum contest and development in Australia. The transformation of Victoria's mathematics curriculum (B. McCrae & Stacey, 1997) emerged from this period, and resulted in Victoria being considered an innovative world leader by the German Werner Blum (1993). Blum (1993) subsequently became an influential member of the PISA mathematics expert group (OECD, 2005b; Turner & Stacey, 2015). Australians with an international influence on curriculum and educational assessment include Peter Fensham (2016) as a founding member of the PISA science expert group (OECD, 2000, p. 103), and Kaye Stacey (2001) as chair of the PISA mathematics expert

group (OECD, 2014, p. 465). More recently, Australians were central to the introduction of technology-based problem-solving assessment in the 2015 cycle of the PISA. This inclusion of technology-based problem-solving resulted from the *Assessment and Teaching of 21st Century Skills (AT21C)* project led by Australians (P. Griffin & Care, 2015; P. Griffin, Care, & McGaw, 2012; OECD, 2013b, 2016b). Australians have therefore influenced the development of several international assessment frameworks.

The emergence of neoliberalism and education markets in Victoria

The emergence of neoliberalism and markets in education have influenced developments in educational assessment in Victoria and Australia more broadly. Similar to other parts of the world, Australian economic policy shifted from a Keynesian (1953/1964) approach in the 1940s towards a monetarist approach (Friedman, 1962/2002) during the 1980s (Laming, 2012; Marginson, 1993). For example, Laming (2012) describes Australia's Commonwealth scholarship scheme established in the 1950s as providing Australian citizens with increased access to universities. This access was further enhanced in the 1970s through the abolition of university tuition fees (Laming, 2012, p. 19). However, the reintroduction of university fees during the 1980s in a style advocated by the economist Friedman (1962/2002) heralded the rise of neoliberalism into Australian education (Laming, 2012; Marginson, 1993; Stedman Jones, 2012).

Educational assessment played a role in shifting control over education from government towards markets, a shift associated with neoliberalism (Marginson, 1993, pp. 71-73). Marginson (1997a) describes the process of marketisation as being attractive to governments because the

introduction of competition strengthened efficiency pressures and management control. Consumers could be used to discipline the work of professionals, with less resistance than when control was exercised bureaucratically. (p. 91)

In the state of Victoria, marketisation and neoliberalism became manifest with the election of the Kennett Government in 1992 (Spaull, 1999). A central reform for Caldwell and Hayward (1998, p. vii), respectively a key advisor and initial minister for education in the Kennett Government, was the decentralisation of school management. Caldwell and Hayward (1998) describe school management in Victoria as already somewhat

decentralised prior to 1992 and initiated a shift in which aspects of education were centralised and decentralised. Caldwell and Hayward (1998, p. 47) decentralised operational autonomy and centralised accountability through new curriculum accompanied by a new assessment regime.

The accountability regime initiated by the Kennett Government included the Learning Assessment Program (LAP) – a Victorian based forerunner to Australia’s NAPLAN (Pearson, 2009). The LAP reported to parents and signalled a shift from teachers being bureaucratically controlled to being held accountable by parents through markets. However, Caldwell and Hayward (1998, p. 91) notably rejected publicly reporting LAP results to avoid “league tables”.

The exposure of Australian education to markets intensified in subsequent years. The LAP was first replaced by the Victorian Achievement Improvement Monitor in 2000, which was subsequently replaced by NAPLAN in 2008 (Pearson, 2009, p. 18). Rudd and Gillard (2008, p. 5), respectively Australia’s prime minister and deputy prime minister at the time, introduced NAPLAN due to concerns for education’s role in economic growth, reflecting the continued influence of human capital theory on Australia’s educational policy. A website called *My School* (myschool.edu.au) was subsequently developed to publicly report each school’s NAPLAN results. Julia Gillard (2010), the federal minister for education at the time of its launch, heralded *My School* as a new era of transparency. However, Mockler (2013), in an analysis of the media reception of the *My School* website, considers the website as government trying to “shirk their responsibilities with regard to education” (p. 7), and to further shift teacher accountability towards markets. In this sense, the *My School* website was consistent with the policy narrative of shifting control of education to markets as identified by Marginson (1997a). The influence of market-based accountability regimes on educational assessment design is explored throughout this study (Caldwell & Hayward, 1998; Hanushek, 2016; Hanushek & Ettema, 2017; Stedman Jones, 2012).

Neoliberalism and social policy

The rise of neoliberalism and markets in education heralded by the election of the Kennett Government in Victoria also affected social policy. Barcan (1996) identifies a range of

interest groups emerging in Australia during the 1970s that sought to influence curriculum during the 1980s, in what was then

a pluralist or multicultural society. These included neo-Marxist radicals, feminists, Aboriginals, ethnic groups, homosexuals, environmentalists ('greens'), the physically and mentally handicapped, and others. (p. 6)

Similarly, Fensham (2016) considers Victoria in the late 1980s as having an education policy climate that was “explicitly responsive to democratic, feminist, and environmental values” (p. 170). However, Caldwell and Hayward (1998), as influential policy actors in the Kennett Government, considered these developments a challenge to authority, and ideas that led to

a weakening of or challenge to authority in a range of social settings, whether it be Vatican II in the Catholic Church, or the protest movement and the Vietnam War, or the rise of feminism, or a more acute sensitivity to the needs and aspirations of the disempowered or minorities. In schools in Australia, especially in Victoria, this meant a challenge to the authority implied in a centralized curriculum. (p. 12)

In response to this perceived challenge to authority, Caldwell and Hayward (1998) centralised curriculum and educational assessment as a way of abating the influence of pluralist approaches to school-based curriculum dominant during the 1980s (Barcan, 1996, 2003).

An argument that the rise of neoliberalism affected pluralist approaches to curriculum in the 1980s and early 1990s is supported by the observations of Stacey (2001) and Fensham (2016). Stacey (2001) laments that by the year 2000, a “bold experiment of assessment driven change and a real focus of problem-solving and modelling in the curriculum” (p. 4) had come to an end in Victoria. Similarly, Fensham (2016) laments that Victoria’s assessment in senior physics was changed in the 1990s as “once more tipping the subject and its assessment back in favour of the boys” (p. 173). Stacey (2001) and Fensham (2016) moved on from this period to become members of respective subject expert groups for the PISA (OECD, 2000, p. 103; 2014, p. 465).

This study identifies two possible explanations for neoliberalism eschewing progressive forces such as feminism in the manner expressed by Caldwell and Hayward (1998, p. 12). One explanation is Friedman's (1962/2008) assumption that takes wants as fixed, which is inherently conservative by ignoring technological and social progress in its attempt to coordinate society through the economy. Further, Friedman (1953/2008), describes his approach to economics as "in principle independent of any particular ethical position or normative judgments" (p. 146). This contrasts to Habermas' (1998) tradition and is a contrast explored throughout this study. A second explanation for neoliberalism eschewing progress is that conservative neoliberal ideological philosophies, such as those expressed by Caldwell and Hayward (1998), are hostile towards progressive forces such as feminism. Only the first explanation related to economic models which are disengaged from normative and ethical consideration is explored by this study.

1.4. This study

Habermas' (1998) pragmatic tradition from which this study approaches its question considers modernity an unfinished project (Habermas, 1997), and finds the ideologies of neoliberalism and postmodernism in the ascendancy (Ball, 2003, 2016a, 2016b; Stedman Jones, 2012). The contemporary pragmatic tradition is therefore sparse making this study necessarily conceptually dense. This section begins by describing the main contribution this study seeks to make, as well as some corollaries that arise from this contribution. An overview of the chapters contained in the remainder of this study is then presented.

Contribution

This study's central contribution is to locate the field of educational assessment within the pragmatic tradition through three conceptual links. First, this study links educational assessment to Habermas' (1983/1996) discourse ethics through Toulmin's (1958/2003) approach to argument. This link is formally addressed in Chapter 5. Toulmin's (1958/2003) approach to argument is used by Kane (1992, 2006) in an explication of validation in assessment, and by Mislavy, Steinberg, and Almond (2003) to structure arguments about test design and use. While Habermas (1981/1985, p. 35) considers Toulmin's (1958/2003) approach useful, he considers it insufficiently normatively anchored and not powerful enough to differentiate between warranted and unwarranted consensual decisions. In response to this limitation, Habermas (1983/1996, 1992/1998)

extends Toulmin's (1958/2003) approach through discourse ethics (Habermas, 1983/1996) and deliberative democracy (Habermas, 1992/1998). This study works towards extending contemporary approach to educational assessment validation through Habermas' (1983/1996, 1992/1998) extended approach to argument.

A second link made by this study relates to educational assessment generating symbolic media through reports, transcripts and other media. This framing is addressed in Chapter 5 that outlines different conceptualisations of symbolic media (Chernilo, 2002; Habermas, 1981/1992; Parsons, 1963a). This framing is elaborated in Chapter 6, which establishes a link between the *performance-sanction paradigm* developed by Parsons and Smelser (1956/2005) and the concept of *intended interpretation* and *actual interpretation* developed by Moss (2016). This link is further supported by the concept of *justification* of universal norms and their *application* in concrete situations developed by Habermas (1994). It is through these links that educational assessment is conceptualised as a form of symbolic media, a conceptualisation useful for informing educational assessment design and use.

The third link relates to the numerical component of educational assessment data becoming uncoupled from its linguistic component. Habermas (1981/1992) uses the term "delinguistified media" (p. 155) to describe media that becomes increasingly decoupled from ordinary language and associates it with "an uncoupling of system and lifeworld" (p. 155). In the field of educational assessment, Wu (2014) identifies this process through a trend of quantitative data becoming more credible than qualitative data in policy-making, an argument repeated in similar forms elsewhere (Gillis et al., 2016; Gorur & Wu, 2015; Wu, 2014, 2016; Wu & Hornsby, 2014). In arguing that statistics alone do not provide hard evidence, and that interpretation relies on personal experience and sense-making, Wu (2014) is thematising Habermas' (1981/1992) pragmatic tradition for the field of educational assessment. The process where the numerical component of educational assessment becomes decoupled from its linguistic component as described by Wu (2014), is taken by this study as an instance of media becoming delinguistified as described by Habermas (1981/1992). This link frames an imperative for maintaining a pragmatic connection between the numerical and linguistic components of educational assessment data.

Corollaries

Several corollaries emerge from this study's central contribution, one being that educational assessment can be explicitly considered a sociocultural activity when it is linked to progress narratives in the technological and social dimensions. Through normatively anchoring educational assessment through discourse ethics (Habermas, 1983/1996) and deliberative democracy (Habermas, 1992/1998), educational assessment also becomes explicitly oriented towards the economy and polity. This orientation is addressed in Chapter 6 where this study explicitly characterises educational assessment as a sociocultural activity, and in Chapter 7 that distinguishes educational assessment from scientific approaches focusing on realism.

In characterising educational assessment as a sociocultural activity, this study engages with a persistent and ongoing tension within the field of educational assessment validity on how science and ethics are addressed. This is a theme developed throughout this study, with an emphasis in Chapter 7.

This study supports the case made by Baird, Andrich, Hopfenbeck, and Stobart (2017) that educational assessment be distinguished from psychological assessment. This case is more fully addressed in Chapter 7. While this study does not pursue how psychological assessment should be conceptualised, this study does elaborate educational assessment validation as an explicit sociocultural activity simultaneously addressing science and ethics.

Through exploring educational assessment as a form of symbolic media, this study finds that educational assessment is not capable of communicating scientifically valid truths (Collis & Rosenblood, 1985; Davies & Goldsmith, 1988; Dunn, 1959, 1961; Stanovich, 1988). This finding emerges from scientific arguments related to multiple comparisons more fully explored in Chapter 8.

A further corollary is that this study identifies two ways of anchoring educational assessment as symbolic media. One emerges from Parsons (1963a) who describes the medium of influence as institutionally anchored, in a similar way to how the PISA is anchored to the OECD (Adams & Wu, 2002). The second emerges from Habermas (1981/1992) who argues that influence cannot be institutionally anchored and is instead anchored in the lifeworld. This second approach is consistent with teacher-led classroom-

based approaches to educational assessment (Cowie & Bell, 1999; Moss, 2003; Shepard, 2006). Both framings therefore apply to educational assessment and the approaches of Parsons (1963a) and Habermas (1981/1992) are explored in a complementary fashion throughout this study.

Differences in system management practices that emerge from educational assessment are also identified. Strategic (Jelinek, 1979; Mintzberg, 1994) and scientific (Taylor, 1911/1998) approaches to management tend to focus on the numerical component of educational assessment. In contrast, administrative (Fayol, 1949/2013) and bureaucratic (Weber, 1946/2009a) approaches are pragmatically focused and amenable to maintaining a connection between the numerical and linguistic components of educational assessment. These approaches to system management are further explored in Chapter 7.

Finally, Habermas (1992/1998) argues that “the only thing that counts is the compelling force of the better argument based on the relevant information” (p. 103), and this study works towards a similar conclusion for how educational assessment *ought* to be designed and used. The overarching narrative that this study addresses is what information and what influences are brought to bear in the course of making the better argument, where the information brought to bear might change overtime in response to progress. In this way, this study is consistent with the argument-based approach to assessment validation developed by Kane (1990, 1992, 2006, 2016a).

This study complements and extends the approach of Kane (2006) to include legitimization concerns in educational assessment validation. Several legitimization tensions are identified and enumerated in Chapter 8, which summarise this study’s response to the question that asks how different attitudes towards progress narratives affect educational assessment design and use.

The structure of this study’s case will now be described.

The structure

The following chapter, Chapter 2, is the literature review that first explores the affordances technology provides to educational assessment. The broader malaise around educational assessment manifest in the academic literature is then further elaborated. The review concludes with some academic imperatives for this study that draws on Flórez

Petour (2015), Moss (2016), Baird and Lee-Kelley (2009) and Elwood (2013). These writers point to gaps in the literature that are consistent with this study's question and motivating issues.

The immanent critique and its use of the of rational reconstruction is described in Chapter 3 which consists of one section. This chapter elaborates on what is entailed in working towards a general theory. The chapter concludes with an enumeration of requirements for rational reconstruction.

Chapter 4 addresses the purposes of education which are framed around a pattern that emerges from Habermas' (1981/1985) work and shared with other traditions. The proliferation of uses for educational assessment data is then addressed, with the final section providing a general approach for educational assessment in relation to the identified pattern of three worlds.

The broader work of Habermas (1976/1979, 1981/1985, 1992, 1981/1992, 1983/1996, 1992/1998) is thematised for educational assessment in Chapter 5. The chapter describes how Habermas (1976/1979) reconstructs the historical materialism of Marx and Engels (1846/2000) from which two uses of educational assessment emerges: one for the system coordination and one for the lifeworld cultural reproduction. These uses are then further developed in an elaboration of different conceptualisations of symbolic media. A dialectic between science and ethics emerges from this reconstruction. The theories of legitimation crises (Habermas, 1975/2005), communicative action (Habermas, 1981/1985), system and lifeworld, colonisation (Habermas, 1981/1992), discourse ethics (Habermas, 1983/1996) and deliberative democracy (Habermas, 1992/1998) are elaborated in the context of educational assessment.

The case that contemporary education, and educational assessment, is oriented towards the economy and polity is made in Chapter 6. The chapter draws on the work of Parsons and Smelser (1956/2005) as well as Habermas (1981/1992). It frames education, and educational assessment, as relating to the economy and polity through the exchanges of wages for labour, goods and services in exchange for demand, and loyalty to norms in exchange for valid norms. The chapter implicates educational assessment in these exchanges. A discussion on how the dynamics in the economy and polity might be coordinated across society leads to the conceptualisation of an *integrative system*. The

chapter describes pragmatism as considering the integrative system a function of government, and neoliberalism a function of markets. Chapter 6 concludes that markets can coordinate societies to relieve the burden on ethical and moral reasoning, but that markets also affect the social structure of society. Three consequences of market-based coordination through educational assessment are then explored to illustrate tensions between educational assessment and its relationship with the economy and polity.

Chapter 7 progresses the case that educational assessment is a sociocultural activity and explores its relationship with science. The chapter has four sections, with the first addressing contemporary debates on whether educational assessment is a sociocultural or scientific activity (Crocker, 1997a; Newton & Baird, 2016a). The second section addresses contemporary ethical issues in educational assessment that emerge from approaches that separate science from ethics in educational assessment, mainly that educational outcomes are caused by Deoxyribonucleic acid (DNA) (Didau, 2016; Shakeshaft et al., 2013). These positions are juxtaposed to contemporary practices in education that challenge the position of Borsboom, Mellenbergh, and van Heerden (2003, 2004) who advocate in favour of the separation of science and ethics in assessment validation (Borsboom & Wijsen, 2016). The third section provides a formal explication of performativity based on the work of Lyotard (1979/1984) and juxtaposes performativity with the *performative attitude* described by Habermas (1981/1985) and Weir (1995). The fourth and final section of Chapter 7 addresses education system management and associates strategic and scientific approaches to system management with the separation of science and ethics. Several tensions are therefore identified between science and the field of educational assessment.

Chapter 8 works towards a reconstruction of educational assessment and towards a general theory that incorporates legitimacy. It begins by exploring the issues of normative validity in educational assessment (Biesta, 2009), and the role of educational assessment in communicating societal values and expectations (Mislevy, Steinberg, & Almond, 2003). A number of issues pertaining to normative validity are then addressed, including processes of justification and application (Habermas, 1994; Rehg, 2011), permanence and transience (Messick, 1989; Moss, 2003), and semiotic resources (Kress, 2003, 2010). How the use of multiple comparisons affects the status of educational assessment as a scientific medium is also addressed, before discussing how the PISA addresses normative

validity. The final section of Chapter 8 summarises some of the central legitimisation tensions identified through this study.

The discussion for this study is provided in Chapter 9, which has three sections. The first section reflects on how the ideological approaches used in this study seek to address complexity arising from technological and social progress. The chapter explores how the ideological positions might be reconcilable through the better argument. The second section revisits the more important contemporary issues for educational assessment identified through this study, mainly contests around: the conceptualisation of constructs; the use of new mathematical models; normative validity in test development; and the reporting of intended interpretations and how this affects stakeholders, particularly students. The third and final section revisits the issues motivating this study.

The concluding chapter, Chapter 10, reflects on the key contributions this study seeks to make in the context of the presented arguments, as well as potential limitations of this study.

Chapter 2. Literature review

This literature review is in three sections. The first section explores progress narratives in educational assessment through the Koomen and Zoanetti (2018) planning and review framework. This section describes key developments in techniques and technological that affect the design and use of educational assessment.

The second section expands on the malaise in educational assessment identified by Masters (2013). This is achieved by exploring contemporary critiques of educational assessment as well as contemporary debates in educational assessment validity.

The third and final section addresses the academic imperative for this study in the context of study question and its focus on progress narratives. Several imperatives are identified, including calls to approach educational assessment reform from a broader perspective (Flórez Petour, 2015), calls for an expanded theory of validity (Moss, 2016), calls for better planning approaches for new educational assessments (Baird & Lee-Kelley, 2009), and calls for an increased ethical focus in educational assessment (Elwood, 2013).

2.1. The affordances of technology

This section addresses the technological progress narrative for educational assessment that emerge from affordances of technology and new divisions of labour (Durkheim, 1933/2012; Levy & Murnane, 2004; Schumpeter, 1942/2008). Binkley et al. (2012) identify technology-based educational assessment as providing two dimensions for change. One dimension relates to business improvement and is described by Binkley et al. (2012) as a *migratory strategy*. The other dimension relates to innovative assessments delivered through technology and is described by Binkley et al. (2012) as a *transformational strategy*. This study considers the migratory strategy as instrumentally focused and addressed through the method of *business process reengineering* developed by Hammer (1990), and Hammer and Champy (1994). Business improvement focuses on instrumental efficiencies. The second dimension of innovative assessment is therefore of greater interest to this study and is explored through the theme of *enhanced semiotic resources* for meaning making, a concept emerging from Kress (2010) who describes a shift from grammar, to semiotic systems, to semiotic resources. The second dimension with its focus on meaning making is of greater interest from a pragmatic perspective. A

brief excursions addressing tensions between the instrumental and ethical dimensions is undertaken before the instrumental affordances of technology are explored.

Intersubjectivity and machine efficiency

The affordances of technology to educational assessment are such that many aspects of educational assessment can be automated giving rise to a tension between human interaction and automation. The efficiencies afforded by technology give rise to critiques such as that of Postman (1993, 1996) and Lyotard (1979/1984) that allude to dystopian scenarios of a world programmed by machines. A dystopian scenario might include students being assessed using machine generated items, invigilated by technology, and responses scored by automated algorithms. This possibility is addressed here through a legitimization tension; a tension between a scenario where students are intersubjectively assessed by teachers, and a scenario where students are assessed through machine processes. The resulting tension between ethics and technology is described here as a tension between intersubjectivity and machine efficiency.

Intersubjectivity is a central concern in Habermas' (1998) pragmatic tradition and the nature of intersubjectivity can change with through technology. Hegel (1807/1977), for example declared that

Self-consciousness exists in and for itself when, and by the fact that, it so exists for another; that is, it exists only in being acknowledged. (p. 111)

A similar sentiment is expressed by Noddings (2010) from a corporeal perspective of care, where

most women writers use words such as relation, response and responsibility, preservative love, attention, attentive love, nurturance, needs, caretaking/caregiving, receptivity, reciprocity, and many words referring to bodies and birth. (p. 146)

Noddings (2010) and Hegel (1807/1977) both address a concern over intersubjectivity which pertains to the relationship between teachers and students in education. The nature of this relationship can be affected by the disruptive affordances of technology (Bower & Christensen, 1995; Schmidt & Cohen, 2010).

The affordances of technology provide the means to replace intersubjective aspects of educational assessment with more efficient machine-driven processes. Processes amenable to automation include item generation (Drasgow, Luecht, & Bennett, 2006, pp. 473-476; Gierl & Lai, 2016; Haladyna & Rodriguez, 2013, pp. 132-151), and essay scoring (Cohen & Wollack, 2006; Shermis, Burstein, Brew, Higgins, & Zechner, 2016). While machines have the potential to be more efficient, they can also reify conceptions of knowledge, remove a sense of narrative and intersubjectivity from student lives, and make errors (Postman, 1993, 1996). These themes are not directly explored by this study and are instead cast as a tension between intersubjectivity and machine efficiency that might be addressed in the pragmatic validation of educational assessment in the context of concrete situations (Kane, 2006).

Technology's affordances

This review frames the affordances of the progress narrative in technology through the educational assessment planning and review framework developed by Koomen and Zoanetti (2018) illustrated in Figure 1 below. The framework consists of eleven nodes addressing conceptual and operational aspects of educational assessment. The framework highlights the interdependency of decisions made within each node. It also emphasises how decisions within each node relate to educational assessment purpose, educational assessment validity, and actions and consequences arising from an assessment. The remainder of this study works towards addressing concerns related to how different attitudes to these new technological affordances affect educational assessment design.

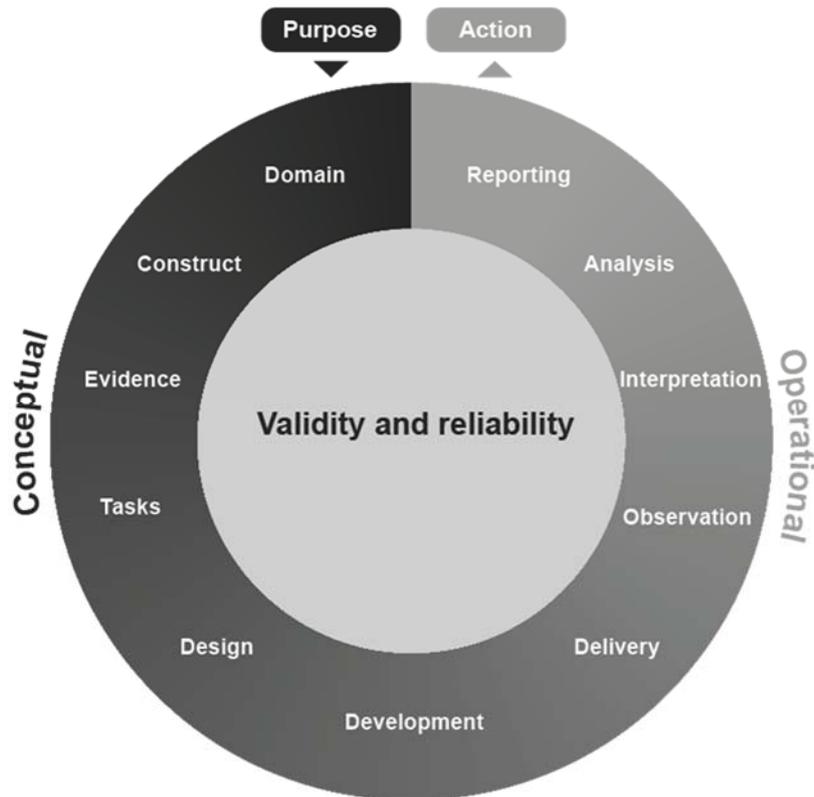


Figure 1 – Planning and review framework for assessment
Koomen & Zoanetti (2018), reproduced by permission © VCAA

Each node of the Koomen and Zoanetti (2018) framework is explored here, with continued emphasis on the construct, evidence, development and reporting nodes throughout this study. These are considered to have more profound effects on educational discourse and are therefore particularly important in design.

The domain node

The domain node addresses substantive information about the target domain of an assessment such as the knowledge, skills and abilities (sometimes referred to as KSAs) used within the target domain. Riconscente, Mislevy, and Corrigan (2016) describe activity in this node as domain analysis and domain modelling, Perie and Huff (2016) describe this in terms of creating content descriptions. Kane (2006), for example, describes the target domain as “representing the full range of possible observations associated with a trait” (p. 33). While the notion of trait has a stronger association with the construct node, there is a strong conceptual relationship between the domain and construct nodes. The domain node has a stronger descriptive and a weaker theoretical

focus than the construct node. Consideration in the domain node may address how a domain is used in society and the economy.

Domain selection in educational assessment can be contested and the assessment of twenty-first century competencies provides an example. Twenty-first century competencies were assessed using technology-based assessment in the PISA 2015 (OECD, 2013b, p. 5; 2016c). This assessment emerged from an Australian project sponsored by Cisco, Intel and Microsoft (P. Griffin & Care, 2015, p. 6; P. Griffin et al., 2012). This development aligns with Lyotard's (1979/1984) critique of performativity where commercial influences define what is valued in education. Ananiadou and Claro (2009, p. 6) share a similar concern arguing that twenty-first century competencies are too economically focused, and their pursuit comes at the expense of the harmonious development of all human abilities. Ananiadou and Claro (2009) observe that twenty-first century competencies may not be relevant to all students, particularly those who will go on to work in economies in developing countries where these skills may not be relevant. This contest over domains points to an ongoing tension between the economy and educational assessment.

The introduction of twenty-first century competencies into the PISA assessment illustrates that broader progress narratives of society are reflected in the selection and the definition of domains addressed by educational assessment. Domain definitions may change as the knowledge skills and abilities used in society change in response in changing modes of production. Changing modes of production in turn create new divisions of labour to create discontinuities and disruption in societies (Bower & Christensen, 1995; Drucker, 1968/1992; Durkheim, 1933/2012; Levy & Murnane, 2004; Marx, 1941/2000; Schmidt & Cohen, 2010; Smith, 1776/1993). This study explores how different attitudes towards these developments are reflected in educational assessment design and use.

The construct node

The theoretical focus of the construct node is reflected by Wilson and Sloane (2000), who describe constructs as “definitions of what students are expected to learn, and a theoretical framework of how that learning is expected to unfold” (p. 183). However, this definition of what occurs within the construct node, and even the term construct itself, is not without

contest. For example, Newton and Shaw (2014, pp. 11-12) consider the term “construct” problematic and prefer the term “attribute”, among other contenders such as “trait” and “disposition”. This dispute about terminology points to a broader theoretical contest involving the construct node.

Borsboom et al. (2003) provide a useful taxonomy for considering the nature of constructs through three ontologies for the latent variable. *Realism* considers the latent variable to exist and to exist independently of measurement, *constructivism* considers the latent variable to be a human construction, and *operationalism* considers the latent variable arising from a “numerical trick” (p. 207) simplifying observations. While Borsboom et al. (2003) argue in favour of realism in assessment validation, their description of constructivism usefully frames the latent variable as a construct.

Borsboom et al. (2003) describe the three ontologies of the latent variable in terms of stances. This is useful because regardless of the ontological argument, the three stances can be used to frame attitudes towards the latent variable. The *realist stance* considers there to be a real referent for the numbers generated through educational measurement, where the referent does not need to be pragmatically communicated (Baird et al., 2017; Moss, 1992). The *constructivist stance* considers the latent variable to have meaning that needs to be linguistically constructed and communicated along with any measurement data for meaning. These stances are useful for describing how educational assessment data is talked about and used across society. Tensions between these two stances is further explored in Chapter 7 and Chapter 8.

A person adopting a realist stance might consider a reported measure to represent a real entity that can be referred to without explication and without reference to educational content. A realist stance is reflected in the analysis of Hanushek and Ettema (2017), who explore productivity in education systems through reported measures without referring to test content. In contrast, a person adopting a constructivist stance might seek to interrogate underpinning assumptions and the content through which reported measures are elicited. A constructivist stance is reflected in an investigation by Wu (2009) into differential country rankings reported by the PISA and the Trends in International Mathematics and Science Study (TIMSS) (Klieme, 2016). Wu (2009, p. 44) accounts for these differences, in part, through factors such as the balance of test content and differential reading load.

The realist and constructivist stances therefore affect the nature of discourse emerging from educational assessment without directly raising ontological concerns.

The construct node represents a central contest in educational assessment, with pragmatic approaches favouring the term construct. Constructivism and pragmatism share a focus on linguistic meaning. From a pragmatic perspective, the construct is socially constructed on the basis that educational assessment addresses a socially constructed relationship with a domain (Reckase, 2017). Borsboom et al. (2003, 2004) consider such a position unsound and instead argue for a form of realism in assessment that requires the latent variable to exist independent of human construction. The term construct therefore reflects a certain stance towards the ontology of the latent variable, a stance that is contested and explored throughout this study with a focus in Chapter 7.

The evidence node

The evidence node addresses the measurement or mathematical model used to analyse student responses (Riconscente et al., 2016, p. 55). The evidence node also addresses the work products students are required to produce for evidence in an assessment. Work products are more appropriately addressed in the context of this study through the development node below. The nature and characteristics of the mathematical models used in educational assessment are therefore addressed here.

Item response theory refers to one set of mathematical models used in many large-scale assessment programs such as the PISA (Adams & Wu, 2002), TIMSS (M. O. Martin & Kelly, 1996), and NAPLAN (ACARA, 2015a). While there are technical differences in how these programs use item response theory, these differences are not pursued here. Instead, item response theory is first described as distinct from, and a development of, classical test theory (de Ayala, 2009; Hambleton, Swaminathan, & Rogers, 1991; Lord & Novick, 1968/2008; Wu, Tam, & Jen, 2017). Classical test theory and item response theory are then explored in the context of other models that are not linear in nature.

The key distinguishing feature of classical test theory (Lord & Novick, 1968/2008) is that the symbolic media the theory produces are test-dependent and test-taker-dependent. As Hambleton et al. (1991) describe, when classical test theory is used, whether “an item is hard or easy depends on the ability of the examinees being measured, and the ability of the examinees depends on whether the items are hard or easy” (p. 7). Hambleton et al.

(1991) argue that this makes it “difficult to compare examinees who take different tests” (p. 7). When classical test theory is used, meaning is made from test scores by referencing questions on the test and students taking the test.

A key distinguishing feature of item response theory is that it has an “explicit latent trait defined in the model” (Wu et al., 2017, p. 21). That is, item response theory is not test-dependent and test-taker-dependent like classical test theory. For example, Rasch (1960/1980, p. xx), the seminal developer of one item response theory model, characterised his project as the ability to make comparisons of students’ ability independent of the instrument used. Wright (1980) hails Rasch’s (1960/1980) model as creating measures that “transcend the questions ... and ... transcends the measuring instrument” (p. ix). Item response theory also produces measures that approach interval and ratio scales of measurement, a development explored by Fischer (1995) following the work of Rasch (1960/1980). The feature of providing student inferences at a ratio level, and where these ratios are not test-dependent, makes item response theory attractive for use in mixed ratios with money such as those used in human capital theory (G. S. Becker, 1964/1993; Hanushek, 2016; Hanushek & Ettema, 2017).

When item response theory is used, meaning is decoupled from test content and is instead anchored on a conceptualisation of a latent variable considered by this study as a construct. Through a shared conception of a latent variable, where deemed appropriate, the numerical scores from one test can be equated and linked to a potentially limitless number of scores from other tests (Dorans et al., 2011; Holland & Dorans, 2006). It is through the process of equating and linking, that educational assessments using item response theory take on a distinctive character from classical test theory. Linking and equating allows for the validity of numerical test scores to be extended along with the validity of the linguistic content of the latent variable. The horizon of linguistic content is explored below in the development node.

The distinction between item response theory and classical test theory is not the only distinction between mathematical models, with the *ruler* and *constellation* metaphors providing another framing. The field of educational assessment, like other fields (Lakoff & Johnson, 1980/2003; Morgan, 1997), often employs metaphor to describe approaches

and methods. The metaphors of the ruler and constellation follow from Mislevy, Steinberg, and Almond (2003), who observe that the

most common way that evidence is accumulated over assessment tasks is summing of item scores, produced task by task, into test scores. Embedding this thinking in formal probability-based reasoning is the basis of classical test theory and its descendent item response theory. (p. 17)

For this form of assessment Wright (1997) invokes the metaphor of “*one ruler for everyone, everywhere, every time*” (p. 42). By way of contrast, Mislevy, Steinberg, and Almond (2003) invoke the metaphor of “unique constellations of knowledge, proficiency, strategies, and tendencies” (p. 22). It is through these invocations that the ruler and constellation metaphors emerge.

The ruler and constellation metaphors are not the only formulation related to this distinction. Reckase (2017) makes a similar distinction between *continuum* and *domain sampling* models to explore differences between psychological and educational perspectives towards assessment.

In the psychology perspective, the goal of instrument development is to produce items or tasks that will estimate a student’s location along a specific continuum. In the educational perspective, the goal is to estimate what proportion of a specific domain has been acquired by students. (p. 1)

Reckase (2017) argues that each model leads to different processes of test development and concludes that the two models are largely incompatible and lead to conflicting uses. While Reckase (2017) provides another formulation for the two metaphors, he does not elaborate on mathematical models.

The ruler and constellation metaphor, or continuum and domain sampling models, are supported by a range of different mathematical models. The ruler model is supported by classical test theory (Lord & Novick, 1968/2008) and item response theory (Hambleton et al., 1991; Wu et al., 2017; Yen & Fitzpatrick, 2006). The constellation metaphor is supported in various ways by Bayesian models (Almond, Mislevy, Steinberg, Yan, & Williamson, 2015), cognitively diagnostic assessments (de la Torre & Minchen, 2014),

diagnostic classification models (Rupp & Templin, 2008), diagnostic measurement (Rupp, Templin, & Henson, 2010), attribute hierarchy method (Leighton, Gierl, & Hunka, 2004), and the rule space method (Tatsuoka, 2009). Further, Braun (2003) observes that

the exponential increase in available computing power and the advent of affordable high speed data networks will affect the design and delivery of tests, lead to novel features and, ultimately, to powerful new assessment systems that are more tightly coupled to instruction. (p. 268)

This is to suggest that the progress narrative of technology will make a wider range of mathematical models, particularly those related to the constellation metaphor, more feasible (Larsson & White, 2014).

The advantage of the ruler metaphor and continuum model is that it is more amenable to comparing and ranking students and makes reporting easier through a single score or related code such as a band, or level (ACARA, 2015a; Adams & Wu, 2002). The advantage of the constellation metaphor and domain sampling model is that it is better at differentiating students in relation to test content to allow for more granular reporting.

Another issue for the evidence node emerging from the narrative of progress in technology is unobtrusive and ubiquitous data collection. This development is extensively addressed by Behrens and DiCerbo (2013, 2014), who also observe a shift from an *item paradigm* to an *activity paradigm*. In the item paradigm students respond to test items and questions, in the activity paradigm students engage with activities such as games that generate data. The type of data generated through the activity paradigm includes stream data from simulation-based and game-based activities (Feng & Heffernan, 2006). Baker and Inventado (2014, p. 63) identify four broad types of models for analysing these data: prediction models, structure discovery, relationship mining, and discovery with models. Behrens and DiCerbo (2014, p. 56) consider this list as not exhaustive, and that the list draws on data mining techniques used outside the field of education. They go on to identify analyses of these types of data as a slowly emerging discipline involving approaches such as learning analytics, data mining, and data science. Chapter 6 further explores unobtrusive and ubiquitous data collection and its problematic relationship to educational assessment and its potential effect on students as citizens.

The tasks node

The tasks node addresses the environment in which test-takers say or do something to produce evidence about the assessed construct, as well as the nature of the work products that need to be produced as evidence (Riconscente et al., 2016, pp. 55-56).

For conventional paper-based educational assessment, testing is generally conducted in a standardised environment. Brennan (2006a, p. 9) argues that while standardisation assists with reliability, the level of standardisation is not a matter of psychometrics, but a normative question for the body politic. This is to suggest that standardisation is more a matter of ethical than scientific reasoning.

The introduction of technology to educational assessment adds new demands on standardisation, and to demands for increased support during testing to present new issues of fairness (Camilli, 2006; McCallin, 2016). Technology also introduces new aspects of fairness for students with disabilities requiring modifications and accommodations (S. N. Elliott & Kettler, 2016). In these ways, the introduction of technology raises new issues of fairness and ethics for consideration in educational assessment design and use.

When educational assessment is delivered through digital technology, questions of fairness and access can manifest at the jurisdictional level. This is illustrated by fifteen countries delivering the PISA 2015 assessment only in paper-based form and not in computer-based form (OECD, 2016b, p. 32). This was presumably due to a lack of technological infrastructure, either nationally or at the school level, given that school infrastructure was used for the technology delivered component of the PISA tests (OECD, 2017e, p. 106). This is to illustrate how issues related to standardisation and fairness can manifest at the student level and at the national level. It also illustrates how notions of fairness can be affected by technological progress.

The infrastructure and standardisation issues raised by technology include, for example, the potential for construct-irrelevant variability due to screen-size (Davis, Strain-Seymour, & Gay, 2013). Computer processing speed and fidelity of visuals is a further issue (AERA/APA/NCME, 2014, p. 65). Some forms of text have been found to be more difficult to comprehend on screen than on paper (Mangen, Walgermo, & Brønnick, 2013). There are other issues related to page navigation, with some studies showing that page-by-page navigation provides for better mental representations of the text compared to

navigation by scrolling (Piolat, Roussey, & Thunin, 1997). There are also domains that require specialised equipment for assessment (Jarvis, Dickie, & Brown, 2013). These are all new issues that technology introduces to educational assessment design and use which relate to standardised environments.

In contrast to standardised environments, others argue for authentic testing environments. Wiggins (1989), for example, argues for contextual realism in educational assessment, where the test environment matches what might be encountered when working in the economy (Newmann, Brandt, & Wiggins, 1998, p. 20). As generations of children become increasingly immersed in a world of technology (Prensky, 2001), what constitutes authentic assessment may similarly evolve, an issue explored by Evidence Centred Design (ECD) through simulation and game-based assessments (Riconscente et al., 2016).

By way of contrast, Bagnato, Goins, Pretti-Frontczak, and Neisworth (2014) argue for authentic assessment in the early years of children's lives but from a different perspective. Bagnato et al. (2014, p. 121) specify natural settings and situations where children are familiar and in familiar routines, such as in a classroom, their home, or in a community setting. The arguments made for authentic assessment environments in the early years by Bagnato et al. (2014) are distinct from the arguments made by Wiggins (1989) for authentic settings. Nevertheless, both serve to illustrate a tension between authentic and standardised environments for educational assessment.

The tasks node also addresses the issue of semiotic resources introduced by technology. Kress (2010, pp. 5-7) argues that technology provides for an increased range of semiotic resources for meaning-making beyond those traditionally provided by text-based grammar. Chomsky (1957/2015), in his seminal study *Syntactic Structures* on linguistics and text-based grammar, explores phonemes, morphemes and transformations for meaning-making. Kress (2010, p. 5) extends this semiotic grammar into a unifying model that incorporates gesture, signing, image, and music in addition to text (Barthes, 1977). Nevertheless, Kress (2003, p. 1) considers that speech will remain the major mode of communication, and writing the preferred mode of communication for the political and cultural elites. This study uses the term "linguistic" in the context of educational assessment to include meaning made through traditional and enhanced semiotic

resources. In this way, Kress (2003, 2010) provides a framing for enhanced meaning-making afforded by technology to educational assessment as described by Binkley et al. (2012).

The concept of semiotic resources relates to item-types used in educational assessment and their validity. Rodriguez (2003) argues that item-types such as multiple-choice, and other selected-response types, have been the mainstay of conventional educational assessment. Sireci and Zenisky (2016), in their review of the literature, identify that technology affords items that can: have multimodal stimuli, accept multimodal responses, be responsive to test-taker input, and interact with other test-takers. Behrens and DiCerbo (2014) address enhanced semiotic resources afforded by technology through an item paradigms and activity paradigm. Scalise and Gifford (2006) provide a taxonomy for describing these items using twenty-eight item-types and seven categories. Potential issues identified by Sireci and Zenisky (2016, pp. 325-326) arising from new technology include that some items may become amenable to test-taking strategies and advantage those familiar with them. The enhanced semiotic resources afforded by technology is therefore accompanied by new validity concerns.

The design node

The design node is concerned with the design of test forms, sometimes described as developing a specification or a blueprint that specifies what types of items or tasks, and how many of each, will make up a test (Downing, 2006). Traditional paper-based test forms are generally of a fixed-form and linear type, with each test-taker in a cohort administered the same test form with items presented in a linear fashion (K. A. Becker & Bergstrom, 2013). This linear form of cohort testing is an aspect of test design particularly affected by affordances of technology.

Downing (2006) considers the complexity of test form design as dependent on the mode of an assessment. Simple paper-and-pencil test forms can be assembled manually, but when multiple or parallel test forms are required the process becomes more complex and may require technological assistance. For example, 66 different test forms were used for the PISA 2015 to ensure coverage of the assessed domains (OECD, 2016b, p. 29). Multiple versions of a test may also be required when an educational assessment is administered to a group, or groups, of test-takers over multiple administrations (Wendler

& Walker, 2016). Tests that are delivered using technology, particularly those that are adaptive, may require specialised software. Adaptive tests increase the complexity of test design to present new issues.

Computer adaptive tests designed to adjust to a student's ability level are becoming more prevalent (Davey, Pitoniak, & Slater, 2016). Adaptive tests are sometimes justified on the basis that it increases student motivation, as motivation can decrease when students are continually presented with material that is either too hard or too easy, and adaptive tests ameliorate this effect (Katz, 2016, p. 475; Wise & Kingsbury, 2016; Wise & Smith, 2016, p. 211).

Improved measurement precision and efficiency is another motivation for adaptive testing. Thompson and Weiss (2011) document procedures for computer adaptive testing including how item calibration, selecting items, scoring responses, and terminating the test can be automated through automated algorithms. These algorithms contrast to, for example, a teacher selecting questions for teacher-developed classroom assessments and can lead to issues of validity and legitimacy. For example, poorly designed algorithms may result in issues related to an inappropriate balance of content over a domain (K. A. Becker & Bergstrom, 2013; van Der Linden, 2016, pp. 525-526). These potential differences between tests developed by algorithms and tests developed by teachers relate to the tension between intersubjectivity and machine efficiency described earlier. Nevertheless, adaptive testing potentially provides more precise estimates in a shorter time (A. J. Martin & Lazendic, 2018; Thompson & Weiss, 2011; Weiss & Kingsbury, 1984).

Australia's NAPLAN program is introducing computer-adaptive testing to both enhance student engagement and measurement precision (A. J. Martin & Lazendic, 2018). However, this transition continues to present comparability issues leading to calls for further investigation (ACARA, 2018; Merlino, 2018; Robinson, 2018). Issues like this that emerge from progress in technology and its use in educational assessment motivates this study.

The development node

The development node addresses the process of test developers and item writers developing test items and questions. The development node is important to this study

because it pertains most directly to the linguistic component of educational assessment, and how this component relates to normative validity (Biesta, 2009, 2015a, 2015b; Braaten, 1991; Habermas, 1981/1985).

The normative focus of test development has a long tradition. Schmeiser and Welch (2006, pp. 324-325) argue that item writers should ideally be educational experts knowledgeable about the content to be assessed, and knowledgeable about the appropriate difficulty levels for targeted cohorts. In the broader normative sense, they also argue that item writers be representative of the target cohort in terms of geography, race, and gender.

The normative validity of an assessment is also affected by how constructs are defined. McCurry (2017) observes that test items

can aim to test knowledge, or they can aim at specific or general skills. They can involve understanding acultural and abstract or on domain-related and contextualised ideas and information. (p. 4)

This illustrates how the cultural and domain specificity of test content can change with the construct definitions and the purpose of a test, which in turn affects normative validity in terms of how the content relates to normative contexts.

McCurry's (2017) historical account suggests that the notion of a construct emerged from a desire to extend the validity horizons of educational assessment. For the Australian CSSE scholarship test implemented from the 1960s, McCurry (2017) describes an initial focus on normative anchoring through loose specifications and the development of content through broad input. McCurry (2017, pp. 78-81) also describes how the notion of a construct emerged out of a desire to conceptually link the CSSE to other examinations. This was achieved through the concepts of construct validity and concurrent validity as well as using numerical techniques such as confirmatory factor analysis. The idea that the construct emerges from test content is supported by Wiliam (2010), who argues that "the construct is made manifest only in the assessment" (p. 259). McCurry's (2017) account suggests that the distinction between construct and test content emerged out of a desire to broaden validity horizons of educational assessments.

Extending the validity horizons of the linguistic component of educational assessment is reflected in methods of equating and linking that extend the horizons of the numerical

component (Dorans et al., 2011; Holland & Dorans, 2006). McCurry's (2017) account on expanding the validity horizons of test material through cross-curricular testing has implications for normative validity. The term cross-curricular in this context, and as used by McCurry (2017), refers to test content that is equally relevant to curriculum across jurisdictions in which a test is administered. McCurry (2017) describes cross-curricular tests as containing "stimulus-based groups of items rather than discrete item testing" (p. 21), where the stimulus material is usually unfamiliar to students and not generally drawn from text books. These stimulus-based items emerged in Australia during the 1960s, and as McCurry (2017) argues, these developments were adopted by the PISA (OECD, 2000). Of interest to the present study is that cross-curricular stimulus-based item testing is somewhat removed from local curriculum to affect normative validity.

The PISA can be used to illustrate that decontextualised cross-curricular testing has implications for normative validity. Grisay (2002), for example, details how linguistic, cultural and curricular bias was addressed in the first cycle of the PISA. While Grisay (2002) reports translation errors as the major cause for items to function poorly in the PISA, cultural and curricular factors were also identified. Issues with translation in test material, as well as issues with cultural and curriculum specificity, become manifest when an item performs inconsistently across countries. For example, the PISA deems an item "dodgy" when it is found to be significantly harder or easier relative to other items for a particular country (Adams & Wu, 2002, p. 105). How consistent an item performs across translations in terms of facility or difficulty, as well as how consistently an item discriminates between students, is determined using linguistic and numerical methods (Adams, 2002; Grisay, de Jong, Gebhardt, Berezner, & Halleux-Monseur, 2007). In relation to issues of "cultural biases or curricular differences" (p. 57), Grisay (2002) describes that equivalency problems "can be avoided somewhat since once the most unstable items have been identified and dropped" (p. 57) a stable scale emerges. Grisay's (2002) account suggests that only test content that is considered equally appropriate across all countries participating in the PISA is selected for the test, with items manifesting cultural or curricular bias dropped. The normative validity of the PISA test content is therefore normative in the sense that it only includes test content similarly applicable to all countries, with unique cultural and curricular elements specific to each

country removed. This is to suggest that as the normative horizon for test content increases, its sociocultural specificity may diminish.

The development node also relates to the tension between intersubjectivity and machine efficiency through automatic item generation (Drasgow et al., 2006, pp. 473-476; Gierl & Lai, 2016; Haladyna & Rodriguez, 2013, pp. 132-151). The efficiency tension is illustrated by Gierl and Lai (2016, pp. 410-411), who identify cost saving as one motivation for automatic item generation. Furthermore, Haladyna and Rodriguez (2013) observe that for many critics, “the main problem with item-writing is that it is a subjective event” (p. 140), with automatic item generation providing the means to remove this human variability to increase reliability. These arguments are not pursued here, other than to draw further attention to the inherent tension between intersubjectivity and machine efficiency.

The development node, as described here, encompasses a variety of issues and contestations that relate to: item-types, semiotic resources, cultural relevance, curricular relevance, and linguistic equivalence. These are important from the perspective of this study in that all numerical evidence from educational assessment used in educational discourse and public policy development are necessarily founded upon, and bound to, test content (Gillis et al., 2016; Gorur & Wu, 2015; Polesel, Rice, & Dulfer, 2013).

The delivery node and observation node

The delivery node and observation node are considered simultaneously here, as activity in these nodes generally occurs simultaneously. Koomen and Zoanetti (2018) differentiate the two nodes to bring attention to the importance of capturing student responses through either fully-functional technology or appropriately trained staff. The issue of capturing student responses through technology, and standardisation of testing environments, relate back to issues described in the tasks node. The delivery and observation nodes bring into focus the role of test supervision or invigilation.

Invigilation is resource intensive and the introduction of technology increases the potential for division of labour through enhanced skills needed to manage and provide specialised instructions and equipment support for technology-based testing (McCallin, 2016, pp. 572-574). Digital devices also open up new possibilities for cheating in an exam and so place additional pressures on invigilators in technology-based testing (Dawson,

2016). Technology-based educational assessment therefore has the potential to increase the demand on resources, as well as make invigilation more efficient through remote invigilation (Milone, Cortese, Balestrieri, & Pittenger, 2017).

The interpretation node

The scoring, marking, and coding of student responses, which is addressed by the interpretation node, is resource intensive and can be automated and therefore relates to the tension between intersubjectivity and machine efficiency. Training of markers can involve extensive training exercises that include ongoing quality assurance procedures (Cohen & Wollack, 2006, pp. 377-378). The PISA, for example, emphasises the importance of appropriate selection, training and management of markers (OECD, 2014, pp. 113-114). Extensive analyses of marker reliability and consistency is also undertaken by the PISA (OECD, 2014, pp. 258-276). The subjective variability and the resource intensity of activities address by the interpretation node generate efficiency pressures.

There have been considerable advances in automatic test scoring with a range of proprietary options available (Cohen & Wollack, 2006; Shermis et al., 2016). These can generate efficiencies as well as lead to performativity where the judgement over what is considered good performance is determined by machine algorithm (Lyotard, 1979/1984). On the one hand, innovations in automatic essay scoring offer increased marking reliability with reduced cost. However, there is also some evidence that automated scoring is not always comparable to human scoring. McCurry (2010) investigated the *National Assessment of Educational Progress* study in the United States (Sandene et al., 2005), and concluded that automated essay scoring was not comparable to human scoring. Discrepancies were particularly evident in tasks that were open and broad. This finding is consistent with advice given by Shermis et al. (2016, pp. 349-350), that not all essay prompts lend themselves to machine scoring. In this way, advances in automated scoring not only relate to intersubjectivity but also to performativity. While automatic scoring may be more efficient and reliable, it can also result in proprietary commercial products defining what is valued through limiting the range of prompts and stimulus for tasks such as essays.

The analysis node

The analysis node is primarily concerned with the implementation of the mathematical model described in detail above in the section addressing the evidence node. The complexity afforded by technology introduces new sources of error that require management. The PISA, for example, undertakes a considerable range of checks to ensure the integrity of the data before it is analysed (OECD, 2014, pp. 166-173). As described in the evidence node, the affordances of technology provide for more complex models and algorithms to be used to make inferences about student responses (Braun, 2003). These are implemented in the analysis node.

The reporting node

It is through reporting that an intended interpretation of an educational assessment result is communicated to other contexts and systems where an actual interpretation may lead to an action or decision (Moss, 2016).

Identity formation is an important aspect of reporting. Consistent with the argument from Moss (2003, p. 16) that teachers make test interpretation obsolete through further teaching, Tierney and Koch (2016, pp. 269-270) identify a desire for privacy among students to facilitate development and learning. They go on to observe that this sense of privacy can be threatened by detrimental curiosity motivated by material gain or power over others, an issue exacerbated by advances in technology. Related to this, the proper and improper use of the digital footprints that students make remains an ongoing area of concern (Berson & Berson, 2006; R. Buchanan, Southgate, Scevak, & Smith, 2018; OECD, 2011). The notion of legitimate and illegitimate use of educational assessment can be tied back to the precepts developed by Parsons (1963a) for institutionalised media, which calls for a normative framework for media of influence to guide appropriate use.

The public reporting of results can also be politically motivated. In the 1990s, the Kennett Government decided to limit reporting of Victoria's state-based LAP to parents and schools to avoid league tables (Caldwell & Hayward, 1998, p. 15). This approach contrasts to a later decision made by the Australian Government to make centralised assessment outcomes publicly available at the school level for reasons of accountability and transparency (Rudd & Gillard, 2008, p. 32). In this case, transparency included the publication of school level results online (Gillard, 2010; Mockler, 2013). This

demonstrates that new forms of reporting are made possible through developments in technology, and that its use can be politically influenced.

The way that educational standards are reported and communicated also affects normative validity. Mislavy, Steinberg, and Almond (2003) observe that educational “assessment communicates values, standards and expectations” (p. 4), and how these are reported can involve detailed methods. Traditional standard setting methods generally rely on panellists to make normative judgements on test material and student responses (Cizek, 2012b; Zieky, 2012). Hambleton and Pitoniak (2006, p. 451) argue that panellists for standard setting programs require appropriate qualifications and expertise, and that the composition of panels should be representative of stakeholders and constituencies. This normative aspect of standard setting regimes is also reflected by Zieky (2012, p. 28), who argues that there is no “correct” or objective cut score able to be determined even in a situation of perfect information and perfect method. Similarly, Baird, Cresswell, and Newton (2000) highlight the need for standard setting bodies to defend and justify educational standards given their value-laden and socially constructed nature. However, the social construction of educational standards has somewhat given way to technical construction with the increasing use of item response theory to affect normative validity.

Item response theory allows standards to be technically or mathematically constructed and in a fashion that removes reference to normative contexts. The PISA, for example, publishes *described proficiency scales* using levels based on mathematically calculated cut-scores. Cut-scores are determined using three numerical variables (Adams & Wu, 2002, pp. 197-216). Once these numerical variables have been agreed to and set, describing scales largely becomes a technical process. Each scale description simply involves a narrative description of items that have an estimated item difficulty within each level. The described proficiency scales for the PISA can therefore be algorithmically and procedurally created once the three variables have been set. This illustrates one method for transforming numerical output into linguistic output. This linguistic output is developed through objective and technical processes however, and contrast to standard setting processes that are normatively determined (Cizek, 2012b; Cizek & Bunch, 2007; Hambleton & Pitoniak, 2006). A similar process is used for Australia’s NAPLAN (ACARA, 2014, p. 78). The growth in use of described scales and algorithmically generated cut-scores illustrates how advances in technology can affect reporting to

support Biesta's (2010) observation that "normative validity is being replaced by technical validity" (p. 13).

The reporting function of educational assessment is particularly affected by the technological progress narrative as it not only effects what information is communicated, such as through described scales, but also how it is communicated (Bower & Christensen, 1995; Karimi & Walter, 2015; Schmidt & Cohen, 2010).

2.2. The malaise in educational assessment, contemporary critiques

This section expands on the malaise, division and disarray in the field of educational assessment identified by Masters (2013), a malaise explored in this study from the perspective of progress narratives.

Critiques of contemporary educational assessment

Many contemporary critiques of education and educational assessment are framed around the concept of performativity, which Lyotard (1979/1984) describes as a concern with efficiency at the expense of concerns for truth, justice and beauty (Ball, 2003). Solomon and Lewin (2016) provide one critique that describes a school's attempt to implement change that is thwarted by performativity emerging from contradictions between concerns for individual student development and system accountability. As a result of this contradiction, the Solomon and Lewin (2016) study concludes that the school was "unable to realise its innovative vision" (p. 237).

From the perspective of Habermas (1981/1992), Solomon and Lewin (2016) are describing an accountability regime where the intended interpretation reported by centralised educational assessment programs are incongruous with activity in local contexts, leading to contradictions between system accountability regimes and school management practices. The system accountability regimes are performative because they define and reward performance in a way that is incongruent to the world experienced by teachers and students. A study by Hardy and Lewis (2016) makes similar findings to those of Solomon and Lewin (2016). Hardy and Lewis (2016) observe effects of performativity in teachers who engage with data "for purposes of compliance but without any real sense of the value of doing so" (p, 1).

In another study expressed through the concept of performativity, Lambert et al. (2015) interviewed fifteen drama teachers and some of their students. These interviews explored the effects of performance regimes on school curriculum and student access to art subjects. The study found that accountability regimes resulted in cuts to art related subjects, resulting in fewer subject choices “with many schools dropping drama and dance” (p. 473). Lambert et al. (2015) conclude by attributing this declining focus on art related subjects to performativity. This conclusion is consistent with a quintessential consequence of performativity described by Lyotard (1979/1984), where technical systems create

a game pertaining not to the true, the just, or the beautiful, etc., but to efficiency: a technical "move" is "good" when it does better and/or expends less energy than another. (p. 44)

This aspect of performativity points to a diminished focus on the beautiful, and hence art, while promoting abstract measurable efficiency outcomes. Lambert et al. (2015) illustrate how educational assessment design can affect curriculum through performativity, a matter that directly pertains to the study question.

Howell (2012, 2017) provides another telling critique of educational assessment that is drawn on throughout this study. The study investigates Australia’s NAPLAN but does not explicitly invoke the notion of performativity. Howell’s (2017) study involved 105 students responding to a simple request to create a drawing on their experiences of NAPLAN and then writing about that drawing. Howell’s (2017) study identified cases of *adaptive* and *maladaptive* responses, a framing used by Lipnevich, Berg, and Smith (2016) and adopted throughout this study. Howell (2017) identified maladaptive responses to NAPLAN,

with some children fearing that they would be judged as foolish or compare negatively with other students. Some children also believed that they would let their family down if they did not perform well in NAPLAN. (p. 583)

These maladaptive responses can be described in terms of the performativity of NAPLAN and its capacity to construct student identity and sense of themselves. The concept of

constructing identities is explored later in this chapter through the performative effects of educational assessment and follows the work of Butler (1988, 1990/2007). Howell's (2017) study noted that student responses were varied however, and that not all students perceived NAPLAN as high-stakes and that not all students had maladaptive responses.

Howell (2017) also conducted interviews with senior staff to make conclusions relevant to this study in terms of educational assessment as a symbolic media of influence (Habermas, 1981/1992; Parsons, 1963a). Howell (2017, p. 572) reports that each school's position in wider society makes it impossible for adults and children to experience NAPLAN in isolation from broader auditing and management discourses that extend beyond institutional boundaries. Howell's (2017) study identified community confusion, differences in how NAPLAN was perceived, and a media narrative of distrust, as affecting how students perceive NAPLAN. Children's perceptions were also influenced by parent perceptions. These findings are important in that they empirically describe how educational assessment, as a symbolic media of influence, can have actual interpretations that substantively differ from intended interpretations.

There are also critiques of large-scale educational assessment such as the PISA that Gorur (2016) casts in terms of performativity. The PISA is considered one of the most influential contemporary global educational assessment programs (Biesta, 2015b; Bonal & Tarabini, 2013; Breakspear, 2012; Grek, 2009). Biesta (2015b) describes the PISA in terms of a seduction of numbers and comparisons, to reflect Lyotard's (1979/1984) notion of performativity's focus on efficiency. Grek (2009) makes a similar critique of the PISA as education being governed by numbers. More recently, an open letter by several academics to the director of the PISA expressed concern over negative consequences from the PISA ("OECD and PISA tests are damaging education worldwide – academics," 2014). The initial legitimacy of the PISA, as well as its contemporary eroding legitimacy evident in critiques describing its performativity, is examined in Chapter 8 to explore how large-scale educational assessment can best meet the needs of economies, societies and citizens.

Validity in assessment

The malaise in educational assessment also manifests in contemporary debates over assessment validity. This debate focuses on science and ethics within the broader pattern of the three worlds described earlier. A special edition of the journal *Educational*

Measurement: Issues and Practice (Crocker, 1997b) focused on the relationship that ethics has with science in assessment validity. In that edition, Crocker (1997a, p. 4) asked whether the consequences of educational assessment use should be part of validation, and if validation should be regarded as a scientific or a sociopolitical activity. A similar question was addressed almost ten years later in a special edition of *Assessment in Education: Principles, Policy & Practice* (Newton & Baird, 2016b).

The tension between science and ethics in educational assessment validity can be traced back to the seminal work of Messick (1989). Messick (1989) uses a matrix in an explication of validity that considers constructs, values and social consequences in a unitary manner. Shepard (2016, p. 275) considers Messick's (1989) matrix a mistake, not by way of its unifying ambitions, but because the matrix segments scientific and ethical considerations in a fashion that invites unwarranted separation.

The tension over how science and ethics are addressed in assessment is persistent in the educational literature and of central interest to this study (Crocker, 1997a, 1997b; Newton & Baird, 2016a, 2016b). Messick (1989) advocates in favour of addressing science and ethics in a unified manner, which is a position more recently supported by Shepard (2016). The unified approach contrasts to those seeking to separate ethics from assessment validation and is argued for by Borsboom et al. (2003, 2004) and more recently by Borsboom and Wijsen (2016). The tension between these two contrasting approaches is specifically addressed in Chapter 7 as it affects how educational assessment is used across society.

Contemporary arguments for isolating ethics from science in assessment validation emerges from the work of Borsboom et al. (2003, 2004) (Borsboom & Wijsen, 2016, 2017). An argument connected to the inclusion of ethics is that of considering social consequences and side effects of using educational assessment data, a consideration that emerges from the work of Messick (1989) (Zumbo & Hubley, 2016).

A tension between science and ethics is also reflected in Habermas' (1998) pragmatic tradition. Habermas (1976/1979) reconstructs the historical materialism of Marx and Engels (1846/2000) to frame societies as developing along two dimensions that have a dialectical tension between them. There is a tension between "progress in objectivating knowledge and in moral-practical insight" (p. 177), or between progress in cognitive-

instrumental rationality and moral-practical rationality. To conceptualise the development of moral-practical reasoning, Habermas (1976/1979, pp. 157-158) draws on Kohlberg's (1971) theories of moral development to describe societies as progressing along a continuum of evolving structures of action, world views, and institutional structures. The inclusion of the moral-practical dimension, or ethics, is the central addition that Habermas (1976/1979) makes to the science focused historical materialism of Marx and Engels (1846/2000).

The validity in assessment literature is extensive and is framed in this study as one between linguistic meaning and numerical levels. The linguistic component emerges from this study's focus on pragmatics (Habermas, 1998; Huang, 2014; Kane, 2013), and numerical component emerges from the field of educational measurement (Brennan, 2006b; Linn, 1989; Thorndike, 1971; Wu et al., 2017). Developments in validity can be described in terms of linguistic meaning and numerical levels, and the associated fields of ethics and science.

The tension between meaning and measurement is reflected in the historical literature of educational measurement. The first edition of *Educational Measurement*, edited by Lindquist (1951), reflected this division with one section devoted to construction of tests that addresses item writing and selection, and another section devoted to measurement theory including reliability and validity. This first edition includes a chapter by Flanagan (1951) that addresses the specific topics of content, ranks, levels of development, growth, and profiles in the context units, scores and norms. In the context of progress narratives, these fundamental elements continue to drive the educational measurement literature, with the argument made in this study that that focus on progress related to test content has diminished in recent years.

The focus on ranks and levels is reflected in the emergence of item response theory and its growing influence (Boomsma, van Duijn, & Snijders, 2001; Embretson & Reise, 2000; Hambleton et al., 1991; Wu et al., 2017; Yen & Fitzpatrick, 2006). The focus on longitudinal growth over time is reflected in an ongoing interest in trend reporting, linking and equating (Dorans et al., 2011; Gebhardt & Adams, 2007; Holland & Dorans, 2006; A. von Davier, 2011). An interest in educational development and growth is also reflected in the contemporary value-added measures to make inferences about students and

teachers (American Educational Research Association, 2015; Harris & Herrington, 2015). An interest in profiles has also emerged from the nomological networks (Cronbach & Meehl, 1955), trait profiles (Kane, 2006), fusion models (Roussos et al., 2007), and attribute hierarchy models (Leighton et al., 2004). However, the linear measures provided by the continuum models and rule metaphor generally dominate the educational assessment validation literature.

Newton and Shaw (2014) extensively detail developments in educational assessment validity. This includes the four main types of validity: predictive, concurrent, content and construct validity (Cronbach & Meehl, 1955). Early formulations of validity identify a tension between content and constructs, and if the two can indeed be separated. This tension continues to manifest in the contemporary context. Wiliam (2010), in one analysis, found “the construct is made manifest only in the assessment” (p.259). Similarly, Baird et al. (2017) found that items define the construct. Different approaches and arguments have put towards these the tension between content and constructs over the decades (Cronbach, 1971; Kane, 1990, 1992, 2006, 2016b; Meehl, 1978; Messick, 1980, 1988, 1989, 1994).

Markus and Borsboom (2013) develop an alternative contemporary approach to validation that focuses on meaning and causation. This approach emerges from the realist stance towards the ontology of the latent trait, where a real entity has the capacity to be measurably influenced by another entity (Borsboom, 2008; Borsboom et al., 2004).

Contemporary approaches to assessment validity emerge from American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, referred to throughout this study as *the Standards* (AERA/APA/NCME, 2014). These standards anchor many validity debates in contemporary practice and academic discourse. The argument-based approach described by Kane (2006) is the current dominant assessment validation paradigm. This approach was more recently recapitulated by Kane (2016a) in a special edition of *Assessment in Education: Principles, Policy & Practice* (Newton & Baird, 2016b). Kane (2006) approaches validation through two arguments, the first involves interpretation of data and its intended use, the second involves the validity of that interpretation. Kane (2006) considers that the interpretive arguments used in validation “cannot be proven” (p. 29).

Instead, interpretive arguments should be judged on their soundness in terms of clarity, coherence and plausibility. In reflecting on developments in assessment validity, Kane (2016a) observes that there is “broad agreement that the evaluation of interpretive claims is at the core of our notion of validity ... [with] less agreement about the role of the consequences of test use” (p. 198). It is the contemporary tension between interpretive claims and social consequences that this study explores through pragmatic framework of Habermas (1981/1985, 1981/1992, 1983/1996, 1992/1998, 1975/2005).

2.3. The academic imperative for a pragmatic approach

This study could not identify a general theory of educational assessment that frames progress narratives across technological, social and personal dimensions. Disparate theories around educational assessment that pertain to the study question do exist. For example, there are theories on how education relates to the economy (G. S. Becker, 1964/1993; Friedman, 1969/2008), theories for educational measurement and numerical methods (Brennan, 2006b; Hambleton et al., 1991; Linn, 1989; Lord & Novick, 1968/2008; Thorndike, 1971), and theories on educational assessment design (Mislevy, Steinberg, & Almond, 2003; Mislevy, Steinberg, Almond, Haertel, & Penuel, 2003; Wilson & Sloane, 2000). There are also related theories, approaches and critiques on the design and implementation of educational assessment and their social impact (Brown & Harris, 2016; Lane, Raymond, & Haladyna, 2016; Pellegrino, Chudowsky, & Glaser, 2001). But no theories were identified that conceptually link all these.

The ECD approach developed by Mislevy, Steinberg, Almond, et al. (2003) and more recently described by Riconscente et al. (2016) is the most comprehensive contemporary theory for technology-based educational assessment design identified. Koomen and Zoanetti (2018, p. 204) identify ECD as an increasingly prominent framework to inform assessment design and draw upon it to develop their framework for planning and review. One limitation of ECD is that it has a distinct performance focus. Both Mislevy, Steinberg, Almond, et al. (2003, p. 2) and Riconscente et al. (2016, p. 40) draw on Messick (1994, p. 16) to describe the constructs ECD seeks to assess, which focus on mainly behaviours and performances. However, Messick (1994) considers behaviours and performances of particular importance only for performance assessment. Messick

(1994) describes performance assessment as distinct from other forms of assessment, and that an

important question is, what constitutes validation for performance assessments? In particular, do performance assessments adhere to general validity standards addressing content, substantive, structural, external generalizability, and consequential aspects of construct validity (Messick, 1989)? Or do they require specialized validity criteria? (p. 13)

Because ECD describes itself in terms of Messick's (1994) performance assessment, and because Messick (1994) himself questions if performance assessment is a subset of possible assessments, this study frames ECD within a broader general theory.

A case for taking a systemic and broader approach towards the issues motivating this study is provided by Flórez Petour (2015). In a study investigating the implementation of *Assessment for Learning* (AfL) in Chile, Flórez Petour (2015) advocates that educational assessment reform be simultaneously approached from historical, systemic, and ideological dimensions. These dimensions respectively address: the long-term and short-term histories within local contexts; the processes of production and circulation within local contexts; and how ideologies affect the application of theories in local contexts. This study's approach is both framed and enabled by the approach taken by Flórez Petour (2015), an approach that is able to engage progress narratives in the context of educational assessment.

A case for broader approaches to educational assessment validation is made by Moss (2016) who describes an ongoing research agenda that addresses local test use. Moss' (2016, p. 248) agenda seeks to examine how different professionals use data, how this data use is supported, and how distinct approaches affect the work of professionals and organisations as well as student learning. Moss' (2016) research agenda, which calls for a more complex theory of validity, is broadly supported through this study by the conceptualisation of educational assessment as a symbolic media of communication (Chernilo, 2002). Moss' (2016) agenda, that frames validity through an intended interpretation in central offices and an actual interpretation in local contexts, provides a way of framing tensions between these two interpretations that emerge from progress in society.

Issues around the nexus between public policy and educational assessment design and implementation is also identified by Baird and Lee-Kelley (2009). Baird and Lee-Kelley (2009) interviewed ten managers in the United Kingdom with responsibilities for implementing new educational assessment programs. Time pressure was identified as a major concern along with issues of complexity and communication. Baird and Lee-Kelley (2009) concluded that

Politically driven reform, combined with the nature of the stakeholder relationships and the scarcity of skilled human resources, conspired to produce a great deal of time pressure on the managers and the process.
(p. 77)

Baird and Lee-Kelley (2009, p. 77) propose developing a generic blueprint for implementing new educational assessments to address environmental complexity, a complexity that is considered by this study as emerging from progress narratives in society. They conceptualise this generic blueprint in terms of a plan clarifying instrumental concerns such as project timelines, as well as addressing social concerns related to community engagement. This study responds to this suggestion by working towards a generalised theory of educational assessment validity that seeks to coherently combine instrumental concerns with political ones in educational assessment design and implementation.

A call for an increased focus on ethics in educational assessment is made by Elwood (2013). In a review, Elwood (2013) concludes that there is a “lack of ethical considerations in large-scale, high-stakes educational assessments” (p. 216). A diminished concern for ethics could be attributed to the influence of neoliberal economics, which Friedman (1953/2008, p. 146) describes as independent of any ethical position, and its effect on educational assessment. This study responds to the concerns over a lack of ethics in educational assessment through the pragmatic approach of Habermas (1983/1996, 1998) which has an explicit focus on ethics.

Closing remarks

This literature review has identified academic imperatives that emerges from a progress narrative in technology in the field of educational assessment. Affordances of technology are affecting developments in all nodes of the Koomen and Zoanetti (2018) design and

review framework for educational assessment. Several issues in education emerging from the use of educational assessment data were identified, particularly issues that emerge from what Lyotard (1979/1984) describes as performativity (Ball, 2003).

The literature review identified continuing tensions in the field of educational assessment validity, particularly around how science and ethics are addressed in validation. While predicative, concurrent, content and construct validity are enduring concerns in educational assessment validity, there is also an enduring tension between content and the construct, and if the construct can indeed be considered distinct from the content.

An academic imperative for a broader approach to educational assessment validity emerges from the academic literature. After Flórez Petour (2015), this study engages with progress narratives to explore reform in a historical context and from the perspectives of different ideologies. After Moss (2016), this study explores expanding the concept of validity to better address local test users and how each might be affected by progress. After Baird and Lee-Kelley (2009), this study explores the relationship that educational assessment has with the economy, the polity, and society more broadly to inform the management of educational assessment design in an increasingly complex work. After Elwood (2013), this study has a particular focus on ethics in educational assessment and the pragmatics of meaning making and communication (Habermas, 1998).

Chapter 3. The methodology

This chapter details the methodology for this study. It consists of one section that first explores immanent critique and critical theory which seek to identify contradictions in society. These forms of critique work towards reconstructing practices in society and the formal method of rational reconstruction is explicated through ten enumerated requirements. This explication draws directly on Habermas (1976/1979). Rational reconstruction provides a more formal basis for immanent critique (Antonio, 1981; Stahl, 2013). Habermas (1981/1985, 1981/1992, 1983/1996, 1992/1998) himself uses this method to reconstruct historical materialism (Marx & Engels, 1846/2000) and to develop theories of communicative action, discourse ethics, and deliberative democracy. These developments are addressed in Chapter 5.

3.1. Immanent critique and rational reconstruction

Habermas' (1998) pragmatic tradition is associated with the methodologies of immanent critique (Stahl, 2013) and critical theory (Antonio, 1981; Kellner, 1992, 2003). Both address practices immanent in society and work towards identifying pathways for emancipatory change. By exploring practices immanent to society, immanent critique is necessarily historically located and takes as its starting point contemporary sociocultural and sociopolitical practices (Antonio, 1981; Stahl, 2013). Stahl (2013) identifies two forms of immanent critique: *practice-based* and *hermeneutic*. Immanent critique can also draw on both forms. The practice-based form draws on the normative practices immanent in society, and the hermeneutic model draws on interpretation of those practices by theorists. This study draws on both practice-based and hermeneutic forms. Similarly, Antonio (1981) describes critical theory in terms of an historically applied logic of analysis addressing issues of the day which is not based on fixed theoretical or empirical assumptions. This study approaches its question through this tradition.

Through being historically located, this study's method and critique is sociopolitically influenced. The method involves a dialectic between historically located fundamental concepts and their sociopolitical manifestation. In this way, decisions over how educational assessment is designed involves both emerging fundamental concepts and deliberative processes. The pragmatic link that immanent critique provides allows the

progress narrative of technology to be addressed in the context of progress in social and cultural dimensions.

This study concurrently employs the methods of immanent critique and rational reconstruction. It employs immanent critique to address contemporary practice, and following Stahl (2013) it draws on practices immanent to society such as the implementation of the PISA and NAPLAN, as well as drawing on interpretations of those practices by theorists. Further, this study draws on interpretations of contemporary educational assessment practices as well as sociological theory, particularly the work of Habermas (1981/1985, 1981/1992, 1983/1996, 1992/1998) and related theorists such as (Parsons & Smelser, 1956/2005). Rational reconstruction is therefore used concurrently with immanent critique to reconstruct contemporary practices and to critique them with emancipatory intent.

The method of rational reconstruction

Habermas (1976/1979) describes rational reconstruction as “taking a theory apart and putting it back together again in a new form to attain more fully the goal it has set for itself” (p. 95). This study extends this ambition to the field of educational assessment, while simultaneously recognising that the goal education sets for itself is socioculturally and sociopolitically influenced. The sociopolitical nature of educational assessment suggests a role for deliberative processes, and this is the focus of following chapters.

Rational reconstruction, like immanent critique, is concerned with reconstructing normative practices within society (Stahl, 2013). Rational reconstruction draws on *formal analysis* that accesses the consciousness of the competent subject. Habermas (1976/1979, pp. 15-16) distinguishes formal analysis from *empirical-analytical* sciences that consider data in an experiential sense. Formal analysis makes essentialist claims requiring a direct correspondence to the structures of reality. The validity of a reconstruction rests on whether it explicates “rules that actually determine the production of surface structures” (p.16). Rational reconstruction “tries to produce an interpretive, fallible, account of practice, but one that does not simply describe what is most obviously visible” (Olson, 2011, p. 142). Rational reconstruction makes essentialist claims, however because it draws on the intuitions of the historically located competent subject, it does not extend to truth and remains open to challenge and discussion (Habermas, 1983/1996, p. 97).

Pretheoretical knowledge

Habermas (1976/1979) describes formal analysis as a “rational reconstruction of concepts, criteria, rules, and schemata” (p. 8) and directed to “domains of pretheoretical knowledge” (p. 14). The planning and review framework of the Koomen and Zoanetti (2018) used in Chapter 2 forms part of the pretheoretical analysis for this study. The Koomen and Zoanetti (2018) framework itself works towards rational reconstruction through making somewhat essentialist claims about all educational assessment. Further, the literature review in Chapter 2 focused on locating the framework in contemporary developments in technology. The Koomen and Zoanetti (2018) framework is pretheoretical as it does not presuppose or describe theories and methods that should apply in each of its nodes. Nevertheless, Koomen and Zoanetti (2018) provide clear case examples of theories and practices that justify the identification of each node. This is consistent with the method of rational reconstruction.

The Koomen and Zoanetti (2018) framework is fallible in that a claim could be made that an activity essential to educational assessment is not included in the framework, or if a node can be shown not to apply to all educational assessment. The formal analysis in this study proceeds on a similar principle.

The Koomen and Zoanetti (2018) framework does not claim the status of general theory however. It has a focus on cognitive-instrumental reasoning and is presaged on external drivers to motivate design. The framework has a focus on what can be done with technology to provide a framing for instrumental considerations pertinent to business process reengineering (Hammer, 1990; Hammer & Champy, 1994). This study seeks to incorporate the Koomen and Zoanetti (2018) framework in working towards a more general theory that incorporates ethical considerations on how educational assessment *ought* to be designed.

Like the Koomen and Zoanetti (2018) framework, the method for this study addresses the pretheoretical basis of educational assessment and does not directly address instances of design and use of educational assessment. However, this study does draw on actual instances such as the PISA and NAPLAN to provide clear case examples where appropriate, nevertheless these are given by way of example and are not the focus of this study. The design and use of educational assessment are amenable to both formal analysis

of fundamental concepts and empirical-analytical investigation of surface structures. Figure 2 illustrates three tiers that emerge from this framing.

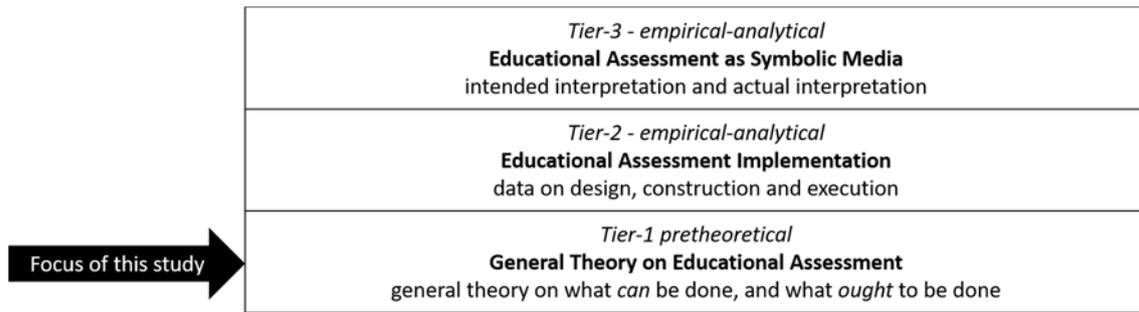


Figure 2 – Three tiers of educational assessment

Tier-1 is the pretheoretical level and object of this study. Tier-2 addresses implementation where the object of empirical analysis might include the quality of implementation such as construct explication, field operations, and methods of reporting. These are described, for example, in technical reports and assessment frameworks produced by the PISA (Adams & Wu, 2002; OECD, 2005b, 2009a, 2012, 2014, 2017e). Tier-2 generally involves discourses of justification in relation to educational assessment design (Habermas, 1994). Tier-3 addresses educational assessment data use and generally involves discourses of application (Habermas, 1994). Tier-3 addresses educational assessment as symbolic media that can be used to coordinate activity including classroom practice through classroom-based assessment and policy development through thematic reports such as those produced by the PISA (OECD, 2016b, 2016c, 2017c, 2017d). The three tiers in Figure 2 illustrate the relationship between general theory and surface structures.

Tier-3 represents the surface structure of educational assessment. The assumptions underpinning the use of an educational assessment can be interrogated through Tier-2, such as how a construct is explicated, how items are developed, which mathematical models are used, and how these relate to broader activities in society. These assumptions in Tier-2 are made manifest and interrogable through pretheoretical general theories identified in Tier-1 as addressed by this study.

The competent subject

Habermas (1976/1979) describes rational reconstruction as a method that reconstructs the practices of the competent subject. This approach contrasts to the transcendental approach taken by Kant (1781/2016) that asserts the existence of fundamental concepts for reasoning that “have *a priori* objective validity” (p. 234). Habermas (1976/1979) uses a weaker transcendental approach where

transcendental investigation must rely on the competence of knowing subjects who judge which experiences may be called coherent experiences in order to analyze this material for general and necessary categorial presuppositions. (p. 21)

Habermas’ (1976/1979, pp. 9-12) work systematically reconstructs the intuitive competence of communicative subjects to develop a theory of communicative action. This study follows in this tradition to reconstruct the intuitive knowledge contained in the practices, theories, and analyses of contemporary educational assessment (Stahl, 2013).

Habermas (1976/1979) distinguishes *description* from *explication*. Habermas (1976/1979, pp. 10-12) correlates description with perceptible and symbolically prestructured reality, and explication with understanding that reality. Description arises from observational experience which is, in principle, a solitary experience. Description typically involves surface structures relating to Tier-3 of Figure 2. Explication is addressed in Tier-1 and Tier-2, which turns away from the world and surface structures and towards understanding “the reality of symbolic formations produced according to rules” (Habermas, 1976/1979, p. 11). In this way, empirical-analytical methods are distinct from rational reconstruction and formal analysis, the first describes surface structures through analysis, the latter pragmatically explicates the intuitions of the competent subject.

Empiricism and rationalism

The distinction Habermas (1976/1979) makes between formal analysis and empirical-analytical methods points to a more fundamental philosophical distinction between *empiricism* and *rationalism*. Guyer and Wood (2016, p. 6) observe that the history of modern philosophy is dominated with a concern over the relationship between what is

perceived empirically and how it is reasoned about rationally. This relationship is often described using the respective terms of empiricism and rationalism.

Empiricism addresses empirical-analysis and involves what Toulmin (1958/2003) describes as *analytic arguments*. Empirical-analytical methods draw on various forms of analysis such as data analysis, syllogistic logic, and semantic analysis of expressions (Habermas, 1983/1996). Toulmin (1958/2003, p. 116) characterises analytic arguments as somewhat tautological as the warrants authorising the conclusions are contained in the conclusion itself. For example, Finland was considered best at reading following the PISA 2000 because it came on top of the reading scale (OECD, 2001, p. 50). This is an analytic claim as the conclusion is contained in the data and can be found through data analysis. This claim is an example of a surface description resulting from empirical-analysis addressed in Tier-3 of Figure 2.

In contrast, rationalism addresses formal analysis and involves what Toulmin (1958/2003) describes as *substantial arguments* where claims require warrants and remain open to rebuttal. Habermas (1983/1996, p. 53) characterises substantial arguments as addressing higher order properties such as what is good and right. For Habermas (1983/1996), substantial arguments draw on the intuitions of the competent subject, and webs of feelings and streams of perceptions that are tested through ongoing discourse (Habermas, 1976/1979, 1983/1996). Rationalism, formal analysis, and substantial argument address how the claim that Finland topped the reading scale is made possible (OECD, 2001, p. 50). Rationalism and explication address arguments in Tier-1 and Tier-2 of Figure 2.

The concern over empiricism and rationalism is associated with a persistent tension in philosophy. Hume (1739/1985), for example, places a greater emphasis on empiricism and perception without recourse to prior or ultimate principle (Garrett, 2015). In contrast, Kant (1781/2016) considers that it “cannot even be seen how there could be a logical principle of rational unity among rules unless a transcendental principle is presupposed” (p. 594). Kant (1781/2016) synthesises the *a priori* and *posteriori* by grounding them in a transcendental unity. Habermas (1983/1996) largely rejects the transcendental approach of Kant (1781/2016) and moves towards a pragmatic approach.

Habermas (1976/1979, p. 12) moves away from Kant (1781/2016) to develop a weak transcendental approach that draws on Ryle's (1949/2000) notion of the competent subject. Ryle (1949/2000) characterises the competent subject as having both *know-how* and *know-that*. Know-how is the ability to produce something and requires an understanding of rules, structures, and their context-specific application. Know-how correlates with description. Ryle (1949/2000) characterises know-that as the competent subject's explicit knowledge and understanding to produce, and a capacity to reveal the "generative structures underlying the production of symbolic formations" (Habermas, 1976/1979, p. 13). Know-that correlates with explication.

Rational reconstruction by the competent subject draws on Ryle's (1949/2000) synthesis of *a priori* and *posteriori* knowledge in a manner distinct from Kant (1781/2016). Ryle (1949/2000) argues that

if, for any operation to be intelligently executed, a prior theoretical operation had first to be performed and performed intelligently, it would be a logical impossibility for anyone ever to break into the circle. (p. 31)

It is through the intuitions of the competent subject that Habermas (1976/1979, p. 12) synthesises and grounds empiricism and rationalism, a grounding that is reflected in both immanent critique and rational reconstruction (Stahl, 2013). This grounding is achieved pragmatically by addressing activity conducted in contemporary society (Bacon, 2012; Huang, 2014).

Rational reconstruction makes essentialist claims that remain fallible. While Habermas (1983/1996, p. 96) considers the intuitions of competent subjects to be generally not fallible, intuitions can nevertheless lead to fallible reconstructions by relying on a wrong choice of example, on a distorted intuition, or by overgeneralising. Reconstructions therefore require corroboration and remain open to challenge (Habermas, 1983/1996, p. 32).

Requirements for rational reconstruction

Habermas (1976/1979) explicates the formal requirement for rational reconstruction that are recapitulated here through ten central points. The method builds on the requirements for explication developed by Carnap (1962). For Carnap (1962), "explication consists in

transforming a given more or less inexact concept into an exact one” (p. 3). The *explicandum* is the less precise concept that is replaced by the *explican*. Carnap (1962, p. 5) provides four requirements for explication that Habermas (1976/1979, p. 13) adopts for rational reconstruction.

1. Explicans should be like the explicandum and should be able to replace the explicandum in all relevant cases.
2. There needs to be exact rules that describe the use of explicans in relation to other scientific concepts.
3. Explicans should be fruitful and useful for formulating general statements.
4. Explicans should be as simple as possible.

Wunderlich (1979, pp. 169-172) critiques Carnap’s (1962) requirements for explication, particularly on the matter of stabilisation of meaning, and calls for the use of clear cases and examples that allow explications to be evaluated with respect to those cases. Consistent with this critique, Habermas (1976/1979, p. 13) adds the following three requirements to the method of rational reconstruction.

5. Explication proceeds with reference to theories, where concepts are explicated to correspond to entire theories, or concepts are explicated interconnectedly.
6. Explication references clear cases so that intuitions can be replaced by exact arguments, reference can also be made to borderline cases that may be explicated separately.
7. Explication uses language that is of the same level as the explicandum and uses ordinary language or a standardised version derived from it.

Habermas (1976/1979, p. 9) emphasises ordinary and uncoded language in explication for formalising assumptions. That explication use ordinary language at the same level as the explicandum precludes the creation of neologisms during rational reconstruction.

Habermas (1976/1979) describes rational reconstruction as leading to general theories. Wunderlich (1979, p. 171), for example, considers the work of Chomsky (1957/2015) and Carnap as leading to general theories. Similarly, Habermas (1976/1979) considers the work of Chomsky (1957/2015) an example of how reconstruction can represent “species competences” (p. 14), and can be compared, in scope and status, to general theories. The final three formal requirements of rational reconstruction emerge from these arguments.

8. Explication addresses the pretheoretical knowledge of competent subjects in the field being explicated.
9. Rational reconstructions are hypothetical and fallible to remain open to challenge.
10. Rational reconstructions, in scope and status, can be compared to general theories.

In summary, rational reconstruction works towards replacing imprecise observations with more precise concepts and theories where explication directly correspond with activities in the lifeworld. This allows common arguments to be replaced by more precise ones.

Closing remarks

This study uses the method of rational reconstruction which builds on contemporary practices in society to work towards a general theory of educational assessment. Rational reconstruction addresses the pretheoretical basic structures of educational assessment by drawing on the practices of the historically located competent subject. Rational approaches, such as rational reconstruction, are distinct from empirical approaches that address interpretation of data generated by educational assessment.

This study’s use of rational reconstruction builds on the know-that of the competent subject, which this study considers in the form of contemporary practice and academic interpretation. Through its focus on the competent subject, the study is both pragmatic which Huang (2014) describes as a systematic study of meaning through language. It is also pragmatic in the way Bacon (2012) describes matter-of-fact problem solving addressing tangible and practical aspects of life. The remainder of this study follows the pragmatic tradition through rational reconstruction and immanent critique.

Chapter 4. Purposes of education and educational assessment

This chapter reconstructs educational purposes which are framed around a pattern emerging in Habermas' (1981/1985) work and shared by other traditions. The purposes are elaborated in reference to immanent contemporary practices in education and interpretation of those practices by theorists. The proliferation of uses for educational assessment data is then addressed, with the final section providing a general approach for educational assessment in relation to the three worlds.

4.1. Pattern of three worlds

The purposes of education and educational assessment will be elaborated here through a pattern of three world that emerges in the literature. *Pattern language* is used in architecture, software engineering, and educational assessment (Alexander et al., 1977; Gamma, Helm, Johnson, & Vlissides, 1995; Seeratan & Mislevy, 2008). Pattern language provides for a coherent picture across paradigms and provides a way for describing and communicating concepts that repeatedly present but never the same way twice (Alexander et al., 1977, p. x; Gamma et al., 1995, p. 2). Pattern language is used here to elaborate the purposes of education.

Habermas' (1981/1985) approach to pragmatics builds on the concept of communicative action that is elaborated in in reference to the objective world, the social world and the subjective world. These three worlds are evident in broader literature on philosophy, sociology, and education. Habermas (1981, p. 8; 1981/1985, p. 176) draws on Weber (1946/2009b) to develop the concept of three worlds. Weber (1946/2009b) identified the autonomous spheres of science, morality and art as becoming increasingly differentiated as religion disintegrated as a unifying metaphysics. Habermas (1981/1985) goes on to elaborate that validity is redeemed with respect to each world through respective claims to propositional truth, normative rightness, and subjective truthfulness. Habermas (1981/1985) describes the rationality used to redeem validity in each world as cognitive-instrumental, moral-practical and aesthetic-expressive respectively. This pattern of three worlds is represented differently by other theorists.

Lyotard's (1979/1984) critique of the progressive narrative in technology and its effect on society references the three worlds through the terms truth, justice and beauty. The

three worlds are also reflected in the work of Kant (1788/2004, 1790/2005, 1781/2016) through the terms pure reason, practical reason, and judgement. Foley (2003) describes the three worlds through the terms physical world, human world and sacred world (West, 2000). Foley’s framing is poignant because it maintains a sense of sacred within the subjective world which Weber (1946/2009b) frames as disintegrating.

The pattern of three world are also reflected in educational discourse. Biesta (2010) refers to the three worlds using the terms qualification, socialisation and subjectification to describe purposes of education. Dewey (1916/2011) similarly refers to labour and vocation, culture, and leisure that are analogous to Biesta’s (2010) framing. Egan (1998) employs the notions of “Plato and truth”, socialisation, and “Rousseau and nature’s guidance” in a similar fashion to describe purposes of education.

The pattern of the three worlds is illustrated in Table 1 under the rubrics of science, ethics and art through which they are generally referred to in this study.

Table 1 – A pattern of three worlds observed through the literature

Discipline	Science	Ethics	Art
Weber (1946/2009b)	science	morality	art
world	objective	social	subjective
validity claim	propositional truth	normative rightness	subjective truthfulness
rationality	cognitive-instrumental	moral-practical	aesthetic-expressive
Liotard (1979/1984)	truth	justice	beauty
Foley (2003)	physical world	human world	sacred world
Biesta (2010)	qualification	socialisation	subjectification
Egan (1998)	Plato and truth	socialisation	Rousseau and nature’s guidance
Kant	pure reason	practical reason	judgement

While Table 1 illustrates a pattern that makes the disciplines of science, ethics and art manageable for this study, the intent is not by way of reductionism as the use of the pattern in concrete situations may require recourse to further argument. The pattern simply provides a way to coherently discuss conceptualisations across different theorists, and some contests within the pattern pertinent to this study are addressed next.

Habermas' neglect of aesthetic-expressive rationality

How each of the three worlds illustrated in Table 1 are considered, characterised and described varies across theorists and is contested. For example, Braaten (1991) and

Kompridis (2006) identify a deficiency in Habermas' conceptualisation that relates the subjective world or aesthetic-expressive rationality. This critique pertains to feminist critique of Habermas' work which is addressed in greater detail later in this section.

A deficiency in Habermas' aesthetic-expressive dimension might be attributable to an asymmetry that emerges when Weber's (1946/2009b) three worlds are juxtaposed to Habermas' (1976/1979) reconstruction of historical materialism. Habermas (1976/1979) reconstructs historical materialism as described by Marx and Engels (1846/2000) to augment its cognitive-instrumental focus with a concern for progress along the moral-practical dimension. In doing so, the reconstruction of historical materialism neglects the aesthetic-expressive dimension that emerges from Habermas' (1981/1985) reconstruction of Weber's (1946/2009b) three worlds. This asymmetry could explain why Braaten (1991, pp. 51-54) considers Habermas less sharply focused on aesthetic-expressive rationality than the other two worlds. It could also explain a critique by Kompridis (2006) addressing art and self-disclosure, that "Habermas' suggestion is nothing more than a stab in the dark" (p. 108). This critique is pertinent as Kompridis (2006) was a student of Habermas. This suggests that while Habermas (1981/1985) identifies and addresses the subjective world, its conceptualisation may be incomplete.

There are two reasons why the subjective and aesthetic worlds are important to this study. The first pertains to the concept of identity which is expressed through aesthetic-expressive rationality. This relates to poststructuralist critiques including feminist critique, and the gap in Habermas' work could explain why various feminist writers have found the need to reinterpret Habermas' work, even if in a largely sympathetic way (Braaten, 1991; Fraser, 1990, 2013; Meehan, 1995). Fraser (1995), for example, identifies a shortfall in Habermas' work but argues that it does not stand in the way of inquiry, but only that it "necessitates that one read the work in question from the standpoint of absence" (p. 22).

The second reason why the aesthetic world is important relates to this study's engagement with Friedman's (1962/2008) "wants as fixed" (p. 13) assumption, and how this assumption relates to progress narratives. A challenge to Friedman's (1962/2008) assumption from the perspective of progress somewhat relies on a dynamic aesthetic-expressive dimension. For example, the relationship between taste and the economy, or

more particularly, the relationship between taste and class structure, was established through the empirical work of Bourdieu (1979/1984).

Habermas' neglect of aesthetic-expressive rationality is not insurmountable for this study. This is because education, particularly in the compulsory years (OECD, 2016a, p. 475), is largely a universal activity through which the aesthetic is addressed from a normative perspective and able to be addressed through moral-practical rationality. However, this study considers aesthetic-expressive rationality, and related concepts such as taste and desire, as dynamic across society and malleable to education. This malleability is assumed to be both in the sense of emancipating self-expression as well as in the sense of denying and suppressing it, with its denial pertinent to poststructuralist critiques related to sexuality, gender and race. This position is supported by the empirical work of Bourdieu (1979/1984), and consistent with an analysis by Meyer (2000), that

Taste in pluralistic society emerges in socially, politically and rhetorically contested institutional spaces where it is negotiated among a plurality of actors ... (p. 33)

That taste is malleable, and subject to public policy, is evident in the position of former Australian politician Barry Jones (1982/1995). Jones (1982/1995), in his book *Sleepers Wake!*, explores the relationship between technology, education, and work. Jones (1982/1995, p. 252) argues that investment in the arts, leisure, craftwork and sport not only promotes personal development but has the potential to stimulate the economy and drive employment. This study proceeds on the basis that education can affect the development of taste to assist or disturb socialisation (Habermas, 1981/1992, p. 141), and that this development can in turn affect consumer demand to shape the economy and public policy (Jones, 1982/1995). The assumption that taste is malleable to education is made throughout this study particularly in relation to Friedman's (1962/2008) assumption that wants are fixed and to explore issues of identity emerging from gender, race and sexuality.

Critique of Habermas conceptualisation of the social world

There are further contestations over how Habermas (1981/1985) conceptualises the social world that pertains to a distinction between universal ethics and personal morality (Aune, 2007; Habermas, 1990, 1994). This critique also emerges from feminist critiques of

Habermas' work. Habermas (1983/1996) draws on Kohlberg (1971) to conceptualise moral development. Kohlberg (1971) himself draws on Piaget (1932/1997). Gilligan (1977), in the feminist tradition, argues that Kohlberg's (1971) theory of moral development does not give adequate expression to the concerns and experiences of women.

Gilligan (1977) develops an alternative sequence of moral development to that of Kohlberg (1971) and argues for the integration of a "feminine voice". Similarly, Pajnik (2006) identifies ambiguity among feminist critiques of Habermas' work, with some arguing for greater openness and inclusivity. The issues raised by feminists includes how abstract universal ethics relate to local contexts in application. Gilligan (1977) observes that Kohlberg (1971) identifies "a strong interpersonal bias in the moral judgments of women" (p. 484), which causes women to be overrepresented in the third stage of a six-stage sequence. Gilligan (1977, p. 486) observes that in general, women are more likely to be aware of context, and seek to develop approaches to morality that avoid hurting others and avoid conflict. Habermas (1994, pp. 153-154) responds to these concerns by arguing that universal norms need to be first declared from the perspective of a legislator to reflect the common interest. Only then, after more abstract norms have been justified, do the unique characteristics of a concrete case calling for the application of regulation come into view. The tension between the different conceptualisations of moral development by Gilligan (1977) and Kohlberg (1971) directly pertains to the design and use of educational assessment.

The tension exposed by Gilligan (1977) relates to Moss' (2016) framing of intended and actual interpretation. Intended interpretation involves the communication of universal norms, and actual interpretation requires sensitivity to contextual factors and the effect on others. This is a tension addressed by Habermas (1994) through the concept of justification and application further described in Chapter 5. The tension exposed by Gilligan (1977) therefore directly pertains to this study's question and shows how gender potentially influences educational assessment design and use.

The social and subjective worlds

There are also contestations over how the relationship between the social and subjective worlds are conceptualised. Habermas (1981/1985) approaches the subjective self as a

socialised and intersubjective self (C. Martin, 2012). Following Habermas (1976/1979), C. Martin (2012) considers that “the self becomes aware of itself as a self through reciprocal communication with others” (p. 56) and is not an atomistic conception of a person. Habermas (1981/1985) refers to the self as relating

itself to itself not by making itself an object but by recognizing in an external object, in an action schema or in a schema of relations, something subjective that has been externalized. (p. 9)

Further Habermas (1992/1998) considers that

autonomy, agency, identity, authenticity and the self to be fundamentally intersubjective phenomena. (Anderson, 2011, p. 91)

This reflexive and interdependent conceptualisation of the relationship between the social and subjective worlds is not shared across all theorists.

Biesta (2009) employs the term subjectification for the subjective world and considers it as “the opposite of the socialization function” (p. 40), and not about insertion of the newcomers into an existing order, but about independence from existing orders. The term subjectification is drawn from the postmodern tradition that has a focus on the subject, and not intersubjectivity (Foucault, 1975/1991). Biesta (2009) therefore conceptualises subjectification, autonomy and authenticity as largely individual matters distinct from socialisation.

The way Habermas (1981/1985) characterises the subjective self as intersubjectively constructed provides a useful framing for considering the self in the institutionalised context of the school as a site for social integration. Considering the self as intersubjectively constructed provides a useful framing, for example, to address the construction of gender as described by Butler (1988, 1990/2007) in a poststructuralist tradition. When construction of the self is considered an activity of the individual, the construction of identity, such as gender construction, is largely considered a responsibility of the individual. When identity is considered intersubjectively constructed, it becomes a reciprocating responsibility of the broader group and of society.

Foley’s (2003) conceptualisation of the subjective world as sacred provides another valuable perspective. Where Weber (1946/2009b) considered religion disintegrating as a

unifying metaphysics, in Foley's (2003) conceptualisation the three-world pattern maintains a place for the sacred and religious life. This too pertains to the relationship with the social world with Durkheim (2001), for example, exploring the religion as a social phenomenon and means for the self to attain emotional security through communal living.

4.2. Purposes for educational assessment

The pattern of three worlds is not always reflected in how educational assessment is used in society with categories such as formative and summative dominating educational discourse. As described in Chapter 1, Masters (2013) argues that adherence to these simple categories can be divisive. Similarly, Newton (2007) argues that

We give the wrong message when we try to simplify assessment purposes by allocating them to a small number of categories (such as formative, summative and evaluative): we imply that the sub-purposes within those categories are importantly alike. (p. 161)

There are now an increasing range of uses for educational assessment beyond this simple characterisation explored next.

Proliferation of educational assessment use

The range of uses for educational assessment has proliferated in recent decades. In 2001, the report *Knowing What Students Know: The Science and Design of Educational Assessment* identified three broad purposes for educational assessment: “to assist learning, to measure individual achievement, and to evaluate programmes” (Pellegrino et al., 2001, p. 2). Six years later, Newton (2007) identified eighteen purposes for educational assessment, among a growing list. Many of these uses were for outside the classroom. Six years later again, the OECD's (2013c, p. 60) report *Synergies for Better Learning* provided a more general framing around three levels of use: classroom, school and system. The proliferation of educational assessment use is consistent with progress narratives where economies are becoming increasingly knowledge-based and dependent on educational outcomes (Drucker, 1968/1992, 1994; Hanushek, 2016; Hanushek & Ettema, 2017; Jones, 1982/1995).

Assessing students is at the heart of the OECD's (2013c, p. 91) three levels, that are illustrated, in an adapted form, below in Figure 3. In the OECD's (2013c) framing, at the first level, educational assessment informs classroom practices. At the second level, outside the classroom and within the school, educational assessment informs teacher appraisal, school leader appraisal, and school evaluation. At the third level of the system, educational assessment informs policy evaluation, program evaluation, and system evaluation. Further, as in the context of Australia (Gonski, 2011), educational assessment can inform funding arguments and decisions. Figure 3 illustrates the importance of student measures for coordinating these activities across education systems.

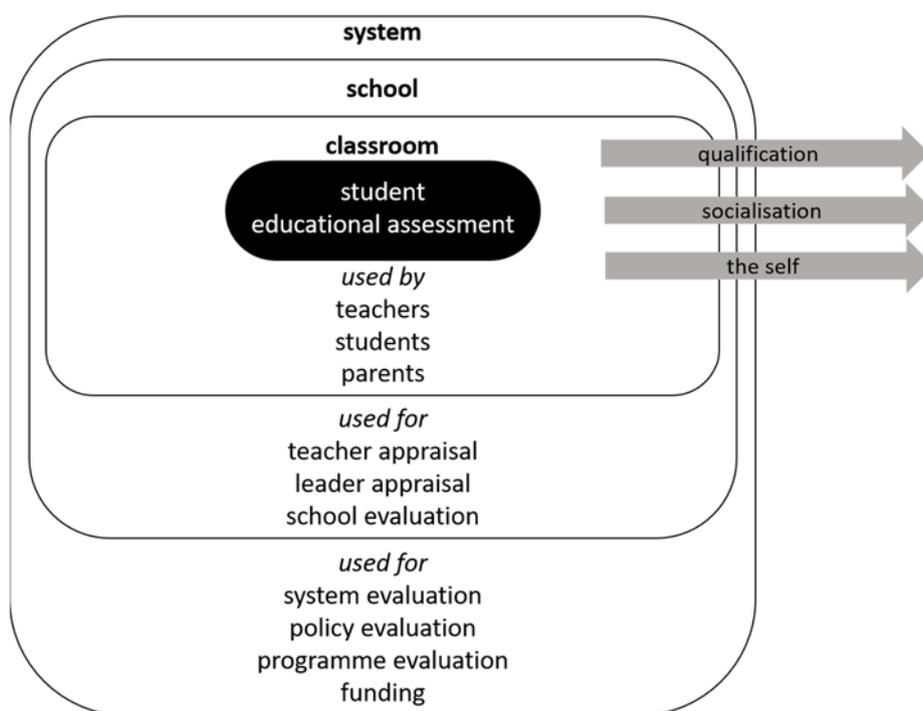


Figure 3 – Levels of use for educational assessment

What Figure 3 also shows, and what is missing in the OECD (2013c) framing, are the three purposes of education as outputs of the education system, or what Oakeshott (1962/1991) refers to as education as an investment in the child. These three purposes reflect the pattern of three worlds shown in Table 1 and described earlier in this chapter.

The terms qualification, socialisation and the self are elaborated in the next section, with their inclusion in Figure 3 highlighting that the OECD (2013c) report generalises *student outcomes* in terms of educational assessment measures and not as investment in the

student. This study frames the purpose of education as investment in the student who emerges as citizen to work in the economy, follows societal norms, and seek to live fulfilled life projects. In failing to differentiate measures of student outcomes from actual investment in students the OECD (2013c) is neglecting the role of educational assessment in the practice of teaching.

This conceptualisation by the OECD (2013c) has two limitations. First, the report does not address how educational assessment benefits students as citizens. Second, it does not describe how educational assessment articulates with systems outside of education such as the economy and polity to conceal the dialectical tension between these systems. These two limitations are addressed throughout this study to provide a better framing of how educational assessment is used across societies to affect the lives of citizens.

The OECD (2013c) identifies a number of drivers for the increased use of educational assessment, with a concern for economic and social challenges faced by countries identified as a central driver. This reflects long held concerns for human capital (G. S. Becker, 1964/1993) and social capital (Coleman & Hoffer, 1987) by governments (Bourdieu, 1986; Marginson, 1993). Other drivers for increased use of educational assessment identified by the OECD (2013c) include increased school autonomy, improvements in technology, and the need for evidence-based decision-making. These drivers are a function of both technological and social progress. These drivers also reflect how ideologies can affect the use of educational assessment, as the focus on evidence-based decision-making reflects an empirical and scientific focus associated with neoliberalism further explored in Chapter 8. The drivers for increased use of educational assessment are therefore multifarious and considered throughout this study as emerging from progress narratives in society.

At the system level, Newton (2007) identifies a “national accounting” purpose for educational assessment. This use of educational assessment is consistent with approaches to human capital theory demonstrated through the work of Hanushek and Ettema (2017) and Hanushek (2013) that respectively use educational assessment to explore productivity in education and economic growth in countries. Australian governments and Australian think tanks also use NAPLAN and the PISA data for advice and planning at a national level (Daley et al., 2016; Gonski, 2011; Lamb, Jackson, Walstab, & Huo, 2015). These

uses suggest that educational assessment is oriented towards activity in the economy and society more generally, a relationship explored in more detail in Chapter 6.

The broadening scope of activities that educational assessment informs creates a tension between classroom use and system accountability use. This tension is identified by Shepard (2000) who observes that

high-stakes accountability testing have reshaped instructional activities to conform to both the content and format of external standardized tests, thereby lowering the complexity and demands of the curriculum and at the same time reducing the credibility of test scores. (p. 3)

This observation shows how system imperatives can reshape classroom activities in a way consistent with Habermas' (1981/1992, p. 325) theory of colonisation of the lifeworld, which is further elaborated in Chapter 5.

Classroom uses of educational assessment

Black and Wiliam (2003, p. 623) describe test use in terms of formative assessment and summative assessment and argue that these labels apply to how an assessment is used and not to the assessment itself. For Black and Wiliam (1998), "assessment becomes formative assessment when the evidence is actually used to adapt the teaching to meet student needs" (p.82). Similar to Shepard (2000), Black and Wiliam (1998) identify issues for educational assessments that are developed externally and administered in the classroom, mainly that externally developed tests

can dominate teachers' work, and, insofar as they encourage drilling to produce right answers to short, out-of-context questions, they can lead teachers to act against their own better judgment about the best ways to develop the learning of their pupils. (p.89)

In this way, externally developed classroom-based educational assessments can arbitrarily impose on classroom practices, and provide a further example on how classroom practice can be colonised by external imperatives (Habermas, 1981/1992, p. 325).

Cowie and Bell (1999) provide a contrasting example of teacher-developed educational assessment. Cowie and Bell (1999) describe an approach that incorporates *student-referenced* and *care-referenced* concerns in addition to *subject-referenced* concerns. In this approach, when a teacher develops a classroom-based educational assessment they might consider student-referenced concerns such as how students understand the assessed subject or domain within the local school context. A teacher might also consider care-referenced concerns that seek to sustain and enhance the quality of interaction among students and between teachers and students. In this way, Cowie and Bell (1999) illustrate an approach towards educational assessment that focuses on the normative lifeworld of students.

Performative effects of educational assessment

Educational assessment can construct identity in a way that is conceptualised in this study through the notion of *performative utterances* otherwise referred to as *performatives*. Austin (1962/1975, p. 6) describes the performative utterance as doing or creating something, such as naming a ship, or saying “I do” in a wedding ceremony. In this way, the performative utterance by doing something is distinct from a constative utterance that describes or declares something to be the case. This concept is important to both Lyotard’s (1979/1984) concept of performativity, and Habermas’ approach to social action, and is further explored in Chapter 5.

Educational assessment results, when reported through transcripts for example, can have a performative effect. In *the Standards* (AERA/APA/NCME, 2014, pp. 216-220), *certification* is considered as demonstrating a level of skill and knowledge in a domain, *credentialing* as signifying an acceptable level of performance, and *licensing* as the granting of legal permission to practice an occupation or profession. It is through certification, credentialing, licensing and qualification that assessment acquires its performative function in Austin’s (1962/1975) sense of doing something to construct identity.

The performative utterance can manifest in several ways. For example, some performatives are *verdictives* that “have an effect, in the law, on ourselves and on others” (Austin, 1962/1975, p. 153). There are also *exercitives* that are given in the form “of a decision in favour of or against a certain course of action, or advocacy of it” (Austin,

1962/1975, p. 154). Importantly, Austin (1962/1975) associates these types of performatives with authority, and pertain to this study's question over how educational assessment relates to institutions including government. A lack of authority can render performative utterances "hollow or void if said by an actor on the stage, or if introduced in a poem, or spoken in soliloquy" (Austin, 1962/1975, p. 22). Therefore, the extent to which educational assessment has the capacity to construct student identity is dependent on the circumstance of reporting, and the institutional authority supporting the utterance.

Important from the perspective of this study, is that the performative utterance, and the construction of identity, is not always tied to institutional authority. For example, Butler (1990/2007) uses the concept of performative effects to describe the constitution of gender, where

acts, gestures, enactments, generally construed, are *performative* in the sense that the essence or identity that they otherwise purport to express are fabrications manufactured and sustained through corporeal signs and other discursive means. That the gendered body is performative suggests that it has no ontological status apart from the various acts which constitute its reality. (p. 136)

In a similar sense, judgements made by teachers through educational assessment, and how they are enacted, can contribute to the construction of student identity.

Following Butler's (1990/2007) conceptualisation of performatives, the construction of identity can occur through classroom-based assessment, where teacher judgements are conveyed to students through verbal or written feedback (Black & Wiliam, 2003; Hattie & Timperley, 2007). Lipnevich et al. (2016, p. 182) identify complex affective and cognitive student responses towards feedback from teachers including adaptive and maladaptive responses. Performative utterances inside the classroom are therefore an important aspect of educational assessment in terms of identity construction and are further explored in Chapter 7.

This study observes that educational assessment can construct student identity through both institutionalised system processes and lifeworld processes. Sadler (2009, p. 159), for

example, considers transcripts and reports issued by institutions have a profound effect on identity and career choices.

The capacity for educational assessment to construct identity brings into question the legitimacy of those constructions. That is, the legitimacy of constructed identities is dependent on the validity of descriptions arising from educational assessment. This distinction is reflected by Austin (1962/1975) who distinguishes between performatives and constatives. Performatives create and are generally either happy or unhappy, while constatives describe and are generally either true or false. However, Austin (1962/1975) considers performatives that have an effect through law (verdictives) to “have obvious connexions with truth and falsity as regards soundness and unsoundness or fairness and unfairness” (p. 152). The legitimacy of how educational assessment constructs identity is addressed in this study by extending the concept of validation (Kane, 2006; Messick, 1989; Newton & Shaw, 2014) and how it meets the needs of society and its citizens in the construction of identity.

Summative assessment

Educational assessment described as summative has two properties of interest to this study. One is that summative assessment can have a performative effect to construct student identity. The second is that when summative assessment is reported it can generate symbolic media through a transcript or other report.

Summative judgements made by teachers and schools reported through grades and the like are a form of symbolic media further explored in Chapter 5. These reports and grades can have value for parents, students and teachers. School reports provide parents with an opportunity to participate in their child’s education. However, Hill et al. (2004) find that existing research does not provide a consistent picture of the parents’ role in education, other than the broad observation that parent involvement in education promotes academic success for their children. Nevertheless, summative reports of student achievement provide a form of media for coordinating the activity of students and their social environment.

Summative assessment is also used for certification purposes, however, Newton (2007) notably omits certification from the list of purposes for educational assessment and argues that

Some might be surprised by the omission of ‘certification’ from this list of purposes. However, ... its use often fails to implicate a specific decision, action or process. (p. 162)

Newton (2007) goes on to observe that certification, which he considers synonymous with grading, is generally used to indicate and testify attainment of a general competence. While certification may have no purpose inside education systems, certificates are symbolic media representing an educational investment in the child (Oakeshott, 1962/1991; Parsons, 1963a). Certificates are useful for the citizen to communicate with other systems of society such as the economy when seeking employment.

Moss’ (2016) concept of intended and actual interpretation is evident when certificates warranting a qualification are used outside of education. For example, a secondary school credential is generally used by students in the transition from compulsory education to either enter employment or on to further studies. When recruiting, employers generally use the qualification “to reduce the initial applicant pool” (Carless, 2007, p. 156). In the actual selection process, employers consider evidence other than the qualification including employment history and interview performance (Ryan, Mcfarland, & Baron, 1999, p. 372). This illustrates a broader principle that normative validity in actual interpretation draws on a broader range of evidence than the intended interpretation described through certification of qualification. This employment scenario also illustrates how Moss’ (2016) conceptualisation is not a novel concept and is instead a reconstruction of practices immanent in society.

Nevertheless, increasing focus on education’s relationship to the economy continues to place pressure on the secondary school exit credential and related symbolic media. Australia’s Chief Scientist, Alan Finkel (2018), in a report motivated by a concern addressing the relationship between education and economic growth, questions the efficacy of the Australian Tertiary Admissions Rank (ATAR). The ATAR is a number rank “between 0 and 99.95 in intervals of 0.05” (VTAC, 2016, p. 2) generated annually by each Australian jurisdiction’s credentialing authority for students completing secondary school. An ATAR can be used across Australia and provides an “overall measure of each student’s performance across all their Year 12 studies” (VTAC, 2016, p.

1). The ATAR is used by universities to assist in tertiary selection. Finkel (2018) is concerned that the ATAR sends unclear signals as it is

used by universities as a tool to simplify their selection processes but perceived, rightly or wrongly, by students, parents and educators in the school system as a driver of subject choices, often towards less challenging and less appropriate subjects. (p. 9)

Finkel's (2018) concerns over the ATAR are shared more broadly.

Blyth (2014, p. 266) identifies issues with the use of the ATAR, while expecting the ATAR to survive in the longer term, albeit alongside alternative methods for tertiary selection. In contrast, Matters (2015) from ACER is less optimistic about the ATAR's future, identifying that it will "progressively break down over the next 15 years or so" (p. 72). Chapter 6 explores these pressures on the ATAR and the effect on the life-projects of students as citizens and the effect of societal coordination more broadly.

4.3. Educational assessment and the pattern of the three worlds

This section works towards reconstructing educational assessment with references the pattern of three worlds and in reference to both classroom and system use. Where institutions such as OECD (2013c, 2016b) generally characterise the outputs of education in terms of measures, educational assessment is reconceptualised here, or perhaps more accurately once again conceptualised, as an activity invested in the child (Oakeshott, 1962/1991).

The stakeholder focus, and the instructional core focus

The student outcome focus illustrated by the OECD (2013c, 2016b), and also advocated by Masters (2005), results in student outcomes being considered in terms of measures for system coordination and not in terms of individual student learning. Considering education in terms of measured outcomes contrasts to approach of Oakeshott (1962/1991, p. 188) who considers education activity invested in the child. Following Oakeshott (1962/1991), the outcome of education might also be considered as the individuated student who has asserted their personal identity through what Habermas (1981/1985) describes as taking a performative attitude towards propositions put the them. In the context of the pattern of three worlds, individuation can also be considered in relation to

the objective, social and subjective worlds. This notion of education being an investment in the student through the student taking a position or attitude is lost when the measures of student outcomes are taken as the product and the outcome of education.

The pattern of the three worlds allows educational assessment to be reconstructed in terms of students taking a performative attitude towards propositions put to them by teachers. Habermas (1981/1992, p. 120) argues that participants in communication simultaneously address the three worlds, while perhaps thematically only stressing one. Similarly, Biesta (2010, p. 5) suggests that the three purposes of education always appear in some mix. This supports the claim that the three worlds are always addressed in communication in the context of education and educational assessment.

That educational assessment can address the performative attitude of students in a way that does not require the generation of data, challenges traditional attitudes towards educational assessment. Newton (2007, pp. 161-162), for example, generally refers to purposes of educational assessment as extracting some form of persistent data for use by third parties, such as system administrators, school principals, parents, credentialing authorities, and so forth. This study instead draws on the field of business process reengineering, which makes no assumptions and

first determines *what* a company must do, then *how* to do it. Reengineering takes nothing for granted. It ignores what *is* and concentrates on what *should be*. (Hammer & Champy, 1994, p. 33)

For education, and following Oakeshott (1962/1991, p. 188), this study considers the fundamental function of education as investing activity in the child, with much of this activity involving the performative attitude of the student.

Following the conceptualisation of the instructional core by Elmore (1996), educational activity becomes invested in the child through the instructional core where teachers and students engage in content. It is through this shared engagement in content that educational assessment attains its fundamental status as

a machine for reasoning about what students know, can do, or have accomplished, based on a handful of things they say, do, or make in particular settings. (Mislevy, Steinberg, & Almond, 2003, p. 4)

In this sense, educational assessment is a machine for reasoning about what a student has accomplished in the context of the instructional core by adopting a performative attitude (Elmore, 1996; Habermas, 1981/1985; Weir, 1995). Furthermore, in this framing feedback received by a student after an educational assessment is an essential part of redeeming validity claims associated with a student taking a performative attitude (Habermas, 1981/1985; Hattie & Timperley, 2007; Lipnevich et al., 2016). Feedback given to students is a discourse of application with respect to universal norms (Rehg, 2011), where this discourse includes justifications that motivate students to adopt those norms (Habermas, 1994). The performative attitude, taken through the quintessential stimulus-response format of educational assessment, represents a fundamental interaction in education, an interaction that occurs in the instructional core and which does not require the generation of data.

This study proposes a reconceptualisation of educational assessment from the conceptualisations between formative and summative assessment. Like Newton (2007) and Masters (2013), this study does not consider these two terms useful in the broader sense. A reconceptualisation is illustrated which shows perpendicular foci of educational assessment. One focus is educational assessment for stakeholders including those inside the classroom such as teachers, as well as those outside the classroom. The second distinct focus is educational assessment for the instructional core through which students individuate to create identity. This reconceptualisation provides a framework through which concepts developed throughout this study can be interrogated.

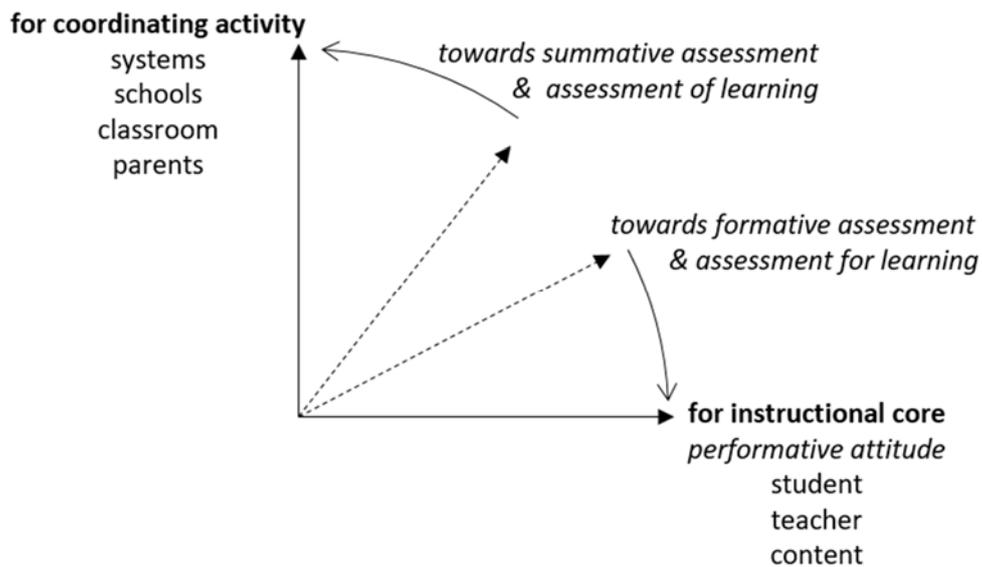


Figure 4 – Perpendicular foci of educational assessment

The instructional core focus primarily addresses the performative attitude of students and describes educational assessment as useful in a way that does not require the generation of data. This focus is largely synonymous with classroom-based educational assessment (Moss, 2003; Shepard, 2000), and is illustrated by the approach of Cowie and Bell (1999) that simultaneously addressed subject-referenced, care-referenced and student-referenced concerns.

The coordination focus of educational assessment is primarily concerned with generating data for stakeholders. For example, that educational assessment such as the PISA and the TIMSS are sample-based educational assessment programs means that students are not the primary beneficiaries of the assessment. This is illustrated by the PISA. In Australia, for example, for the PISA 2015 there were 276,072 students eligible because of their age to participate in the study, however only 14,530 students, or around 5.3% of eligible students, were sampled and participated (OECD, 2017e, p. 205). That is, for the PISA only a small percentage of students are provided with the opportunity to engage with the quality of the PISA test materials through the performative attitude. Furthermore, given that the PISA does not provide student feedback, there is no opportunity for students to redeem the validity of the responses they provide. In this way, large-scale educational assessments such as the PISA illustrate the steering and stakeholder focus of educational assessment for purposes outside of the instructional core (Elmore, 2016; Newton, 2007).

The two perpendicular foci illustrated in Figure 4 can also be used to illustrate educational assessments that address both foci. Where the PISA is totally focused on stakeholders, and teacher-developed classroom-based assessment totally focused on the instructional core, some assessments can address both. For example, the NAPLAN is a population-based assessment that provides both information for stakeholders and also allows students to engage in the performative attitude (ACARA, 2015a). Wyatt-Smith, Klenowski, and Gunn (2010) provide another interesting example of educational assessment where teacher judgements and interpretations are reconciled with system expectations through a process of moderation. This suggests another form of educational assessment that address both foci.

Figure 4 is not a comprehensive illustration of all educational assessment, nor does it directly reveal all tensions in educational assessment. Shepard (2000), for example, identifies high-stakes accountability testing as reshaping activities in the instructional core in a manner consistent with how Habermas (1981/1992, p. 325) conceptualises colonisation of the lifeworld. Black and Wiliam (1998) identify similar issues for externally developed educational assessment. Figure 4 does not directly illustrate this form of colonisation but does assist in its explanation as colonisation involves the two foci being at cross-purposes. Hardy and Lewis (2016) explain instances of these cross purposes in terms of contradictory logics arising as tension from system coordination imperatives being incompatible with imperatives that arise in the lifeworld of the classroom and the instructional core. Nevertheless, despite limitations, Figure 4 does provide a framing for assessments that address both foci in a way that may be complementary, or which may give rise to contradictory logics.

The conceptualisation in Figure 4 suggests that some forms of educational assessment are silent to external stakeholders, yet nevertheless important to the practice of education. Black and Wiliam (1998) describe educational assessment as concerned with “measuring the outputs from the black box of the classroom” (p. 88). However, the classroom only appears as a black box to those outside of it, while those inside of it may be experiencing purposeful activity and individuating through the performative attitude. Tierney and Koch (2016, pp. 269-270) highlight the importance of privacy to students and observe that privacy can be threatened by curiosity and the desire for material gain over others. This desire for privacy leads to a concern over digital footprints that students make (Berson &

Berson, 2006; R. Buchanan et al., 2018; OECD, 2011). In light of the observation by Moss (2003) that a classroom teacher's role is to make test score interpretations obsolete through further teaching, some forms of educational assessment may be best left silent to external stakeholders. How the three purposes of education relate to educational assessment is addressed next.

The three purposes

This study's conceptualisation of education's three broad purposes draws on various traditions described earlier in this chapter (Biesta, 2010; Egan, 1998; Foley, 2003; Kant, 1788/2004, 1790/2005, 1781/2016; Lyotard, 1979/1984; Weber, 1946/2009b; West, 2000). Egan (1998), for example, in *The Educated Mind*, considers that in

the case of the modern school, three distinctive aims have attended its development. It is expected to serve as a significant agency in socializing the young, to teach particular forms of knowledge that will bring about a realistic and rational view of the world, and to help realize the unique potential of each child. (p. 10)

Egan (1998, pp. 10-17) goes on to describe these under the rubrics of: socialisation, Plato and the truth about reality, and Rousseau and nature's guidance (Rousseau, 1762/1979). Egan (1998) proposes an integration of these three aims, as well as proposing a new idea, drawing on Vygotsky (1978), to connect "cultural development in the past and educational development in present" (p. 267). Egan's (1998) fourth new idea, which he himself admits as not being "exactly gleaming fresh from the mint" (p. 26), is addressed in this study through the notion of progress narratives. So, while Egan's (1998) idea of progress is not new, it is consistent with the pragmatic approach to progress adopted by this study.

The three purposes of education directly relate to how education prepares students for work and life. These connections are further explored in Chapter 6 and are described here through qualification that relates to work in the economy, socialisation related to the adoption of societal norms, and the self in terms of living fulfilled lives.

Qualification

The term qualification is used here to address the objective world and draws on the conceptualisation of the three worlds by Biesta (2009). Qualification also reflects Dewey's (1916/2011) interest in labor and vocation (Weber, 1946/2009b). However, qualification does not represent or resemble an objective, physical, or scientific entity. Instead, a qualification refers to a social construction. Qualification is also related to transcripts and certification, and depending on the circumstances in which educational assessment is reported (Austin, 1962/1975), a qualification can take on the characteristic of a social fact, or be *factual* to have *facticity*. It is the facticity of the qualification that provides it with its objective characteristic, where the qualification is a factual artefact of activity invested in the child. The concept of the facticity was developed by Heidegger (1996, p. 52) to express aspects of human existence that exist in a concrete and worldly sense but in a way that contrasts to the factual nature of inanimate existence (Moran, 2000, p. 223). The factual and facticity, and its relationship to validity and legitimacy, is the focus of Habermas' (1992/1998) work *Between Facts and Norms*. Zum (2011) describes the *factual* as

generated, administered and enforced by actual social institutions beyond the control of any individual... (p. 162)

In this sense, qualifications become objective facts when issued through legally authorised institutions.

Qualification in this context however refers to a purpose of education that is present throughout schooling through activities invested in the child (Oakeshott, 1962/1991, p. 188) that are oriented towards the work in the economy and life in society. That is, not all activity related to qualification results in a qualification in terms of a certificate. Biesta (2015b) associates the qualification purpose of education with "knowledge, skills and understanding" (p. 40), and with a student's disposition to do something. In a similar sense, Riconscente et al. (2016) use the terms of knowledge, skills and abilities in their explication of ECD, terms drawn from the human resource management literature (Prager, Moran, & Sanchez, 1997; Stevens & Champion, 1994). This study associates all these descriptions with the qualification purpose of education and considers these present

from the start of schooling, with perhaps increasing emphasis towards the end of the compulsory years.

The purpose of qualification, as conceptualised here, does not refer to the process of certification, accreditation and licensing that might occur in an employment context, or be the subject of a particular educational assessment event. Instead, this study considers qualification as referring to activities that occur in schools and classrooms that are oriented towards the economy and work. It is through its relationship to work and the economy that gives the qualification purpose a certain primacy in contemporary educational assessment, with purposes related to socialisation becoming increasingly important in knowledge economies (Drucker, 1968/1992).

Socialisation

The term socialisation is used here to represent educational purpose with respect to Habermas' (1981/1985) social world and is a term with an extensive heritage in sociology. Where qualification refers to a purpose oriented towards facticity and law, socialisation is normatively anchored. Socialisation happens inside the black box of the classroom (Black & Wiliam, 1998), and occurs in the medium of ordinary language (Habermas, 1981/1985). In this way, socialisation is largely silent to external stakeholders in terms of data, with failures in socialisation more likely to be noticed outside the classroom. Habermas (1981/1992, pp. 141-143) associates failure of socialisation with psychopathologies such as alienation, rupture of traditions, and a withdrawal of motivation (Bowman, McKinstry, & McGorry, 2017; McGorry, Bates, & Birchwood, 2013; Rickwood, Telford, Parker, Tanti, & McGorry, 2014). Socialisation addresses how individuals relate to peers, family and the broader sociocultural environment. Biesta (2009) describes socialisation as concerning how “we become members of and part of particular social, cultural and political ‘orders’” (p. 40).

From the perspective of the student, socialisation begins with

the construction of the social world that the socialized youth finally encounters as the normative reality of the generalized other. In learning how to follow norms and to take on more and more roles, he acquires the generalized ability to take part in normatively regulated interactions. (Habermas, 1981/1992, p. 40)

Further, the processes of socialisation are at the same time one of individuation (Habermas, 1981/1992, p. 58), and has both an affective and moral component (Parsons & Platt, 1970).

Social integration can also be considered as part of the socialisation purpose of education, even though it is sometimes conceptualised distinct from socialisation (Mouzelis, 1992, 1997). In this sense, social integration through schooling encompasses how different social groups and cultures socially integrate, and addresses social integration on attributes such as gender, sex, race and ability (Butler, 1988; Foucault, 1984/1988, 1976/1990, 1984/1990; Said, 1978/1994; Spivak, 1985/2010; Thurlow & Kopriva, 2015). The role of education and educational assessment in socialisation is further addressed in Chapter 5.

The self

The term of the self is used here to represent educational purpose with respect to Habermas' (1981/1985) subjective and aesthetic world. The notion of the self is a broad term that encompasses various traditions. Zahavi (2005), for example, associates the self with "experience, self-awareness, and selfhood" (p. 2). Csikszentmihalyi (1990/2008) describes the self as containing everything that has passed through consciousness: "all the memories, actions, desires, pleasures, and pains are included in it" (p. 34). This description by Csikszentmihalyi (1990/2008) reinforces the observation by Oakeshott (1962/1991, p. 188) that education is activity invested in the self of a child, even activity related to objective and social worlds.

The self, and how it is addressed in education, is becoming an area of interest to the OECD. Kautz, Heckman, Diris, ter Weel, and Borghans (2014), in an OECD working paper, describe attributes that pertain to the self as non-cognitive skills that include personality traits, goals, character and motivations. Kautz et al. (2014) conceptualise non-cognitive skills in terms of a five-factor model that includes: Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness (Costa & McCrae, 1993; Digman, 1990; R. R. McCrae & Costa, 2004). Farrington et al. (2012) similarly hypothesise a model of five factors that includes academic mindset, social skills, academic perseverance, learning strategies, academic behaviours, and academic performance. These are all attributes of the self, which are expressed in relation to both social and objective worlds.

Communicative action and the purposes of education

The broad purposes of qualification, socialisation and the self, as described here, are proffered by way of rational reconstruction of patterns described by competent subjects. The terms themselves, and their contents, remain the object of contestation and the better argument in concrete situations. Further, following Habermas (1981/1985), a challenge related to any one purpose always raises validity claims with respect to all three worlds, even if only one purpose is stressed. For example, a student might ask a question about what qualification to pursue, but such a question not only raises a qualification question of the objective world, it may also raise a challenge with respect to the social world as to what qualification is normatively appropriate. Such a question may also make a claim addressing the student's own subjective world of capabilities and desires. In this sense, communicative action addressing the three worlds are involved in any validity claim regarding any one purpose of education.

The three purposes of education as identified here is important for ensuring that appropriate emphasis is given to each world in educational assessment design and use. Appropriate emphasis fosters a student's ability to: participate in the economy through work, participate in society through agreed norms of behaviour, and develop a strong sense of self and life-project. These purposes are important to both the citizen and society.

Closing remarks

This chapter has elaborated broad purposes for educational assessment with respect to a pattern of three worlds found in the work of Habermas (1981/1985) and the work of others. Braaten (1991) finds deficiencies in Habermas' conceptualisation of the aesthetic world, a deficiency Fraser (1995) considers as not sufficient to reject Habermas' approach. Gilligan (1977) also provides a challenge to Habermas' elaboration of the moral development, but again this contest doesn't provide an insurmountable obstacle for using Habermas' work when addressing educational assessment.

Biesta (2010) and Foley (2003) provide alternative approaches for the three-world pattern that distinctively frame the social and subjective worlds. Biesta (2010) considers subjectification as turning away from socialisation. The sacred world described by Foley (2003) provides a place for religious life that is considered as disintegrating in the framing

of Weber (1946/2009b). Both alternatives take the pattern into trajectory distinct to that of Habermas (1981/1985).

Following Oakeshott (1962/1991), this study considers education as an investment in child in the context of the three worlds. Educational assessment has been conceptualised here not only as a process for generating data allowing for system coordination, but also as a means for a student to individuate with respect to the three worlds through a performative attitude. The following chapters continue to explore these themes.

Chapter 5. Educational assessment for systems and lifeworlds

This chapter thematises the work of Habermas (1976/1979, 1981/1985, 1992, 1981/1992, 1983/1996, 1992/1998) for educational assessment. Habermas is contemporary to this study and draws on a tradition some describe as going back to Kant in the eighteenth-century (Bowie, 2003; de Berg & Large, 2012).

The first section describes Habermas' (1976/1979) reconstruction of historical materialism developed by Marx and Engels (1846/2000). Several concepts emerge from this reconstruction including a focus on progress along the moral-practical dimension in addition to the cognitive-instrumental dimension. Other concepts emerging from this reconstruction include a theory of system and lifeworld, as well as a theory of legitimation crises. The second section elaborates the concept of symbolic media and draws on Chernilo (2002) who conceptualises distinct framings by Habermas (1981/1992), Parsons (1963a, 1963b) and Luhmann (1992, 1984/1995).

The third section details colonisation of the lifeworld that directly pertains to the critique of educational assessment detailed in the literature review in Chapter 2. The fourth section goes on to describe a theory of social action and links this theory to educational assessment through the concept of the performative utterance (Austin, 1962/1975). The concept of the performative utterance is also explored in terms of intended interpretation and actual interpretation described by Moss (2016).

The fifth section addresses a central contribution of this study which is to link Habermas (1983/1996), through Toulmin (1958/2003), to validation in educational assessment as described by Kane (2006) and Mislavy, Steinberg, and Almond (2003). This is achieved through discourse ethics which addresses interpersonal communication that is institutionalised through deliberative democracy.

To frame how legitimacy concerns might be addressed in validity arguments, deliberative democracy (Habermas, 1992/1998) is described in the sixth section with reference to processes of justification and application (Habermas, 1994). It also describes educational assessment's relationship to law. The seventh and final section is a rejoinder to postmodernism to address some of the questions that it poses for pragmatic approaches to educational assessment design.

5.1. From historical materialism to system and lifeworld

Habermas (1981) locates his work in the “Enlightenment tradition” (p. 10) of modernity and makes a significant contribution to it by conceptualising crisis tendencies. Where Marx’s (1859/2000, p. 425) critique envisaged social revolution, Habermas (1975/2005) sought to incorporate what he considered Marx’s (1859/2000) neglected sociocultural dimension through a “theory of communicative competence” (p. xxv). In doing so, Habermas (1975/2005) augmented historical materialism with a progress narrative in the moral-practical dimension as well as the cognitive-instrumental one. While Habermas neglects the aesthetic-expressive dimensions relating to taste as described in Chapter 4, it is the cognitive-instrumental dimension that frames technological progress, and the moral-practical dimension that frames social progress. Habermas (1975/2005) reconstruction replaces Marx’s (1859/2000) theory of revolution with a theory where crisis tendencies are addressed through ethics.

Habermas’ (1976/1979, 1981/1992) reconstruction of historical materialism reframes Marx’s (1859/2000) concepts of *superstructure* and *base* with system and lifeworld perspectives. Marx (1859/2000) developed the concept of superstructure to describe how “men enter into definite relations that are indispensable and independent of their will” (p. 424). Marx (1859/2000) describes the superstructure as supporting society’s modes of production to shape its social consciousness. Habermas (1976/1979, pp. 144-145) considered Marx’s (1859/2000) formulation too focused on technology and economics and developed universal pragmatics as a way to separate communication in the lifeworld from economic exchange. Habermas’ (1976/1979) reconstruction of Marx and Engels (1846/2000) provides the basis for Habermas’ (1981/1992) system and lifeworld theory.

System and lifeworld perspectives

Habermas’ (1981/1992) system and lifeworld model provides two perspectives each drawing on distinct intellectual traditions. The system perspective draws on the tradition of superstructure (Marx, 1859/2000) and functional systems (Luhmann, 2016; Parsons, 1937/1968a, 1937/1968b, 1951/2012; Parsons & Smelser, 1956/2005). The lifeworld perspective draws on theories of action (Parsons & Shils, 1951/1962; Weber, 1915/1964) and the phenomenological lifeworld elaborated by Husserl (1954/1970). These two orientations respectively reflect a focus on system coordination and the classroom in the

context of educational assessment. The dynamics between system and lifeworld perspectives are used by Habermas (1981/1992) to elaborate crisis tendencies in society.

Habermas' (1981/1992) system perspective emerges from differentiation through "the media of money and power, the subsystems of the economy and the state" (Habermas, 1981/1992, p. 307), which shear off from mutual understandings of the lifeworld. The system and lifeworld model also distinguishes between labour and social interaction, and supplement Marx's (1859/2000) concern for economic analysis to include analyses of culture in the lifeworld (McCarthy, 1975/2005, p. xxii). Habermas (1981/1992, 1975/2005) abandons a class-based approach to crises developed by Marx (2000) and replaces it with pathologies that arise from system perspectives colonising the lifeworld (Heath, 2011, p. 74).

Habermas' (1981/1992) lifeworld perspective emerges from Husserl (1954/1970), who conceptualised "a sense-world, a world of sense-intuition, a sensible world of appearances" (p. 106) where living bodies and consciousness play a role. Husserl (1954/1970) considered the lifeworld as a world phenomenon correlative of intentional pretheoretical experience inserted between the world of nature and the world of culture (Moran, 2000, p. 181). Habermas (1981/1992) develops the concept of the lifeworld as a "concept complementary to that of 'communicative action', and as the context-forming background process of reaching understanding" (p. 204). In the context of educational assessment, the lifeworld perspective focuses on the classroom and immediate sense-world and experiences of students and teachers.

The two perspectives of system and lifeworld provide the means for Habermas (1981/1992, p. 150) to distinguish between *social integration* and *system integration* which also derive from distinct theoretical traditions. This distinction again also pertains to the two foci of educational assessment elaborated in Chapter 4, with social integration corresponding to educational assessment for the lifeworld of the classroom, and system integration for system coordination.

Social integration

Habermas (1976/1979) establishes a "connection between system integration and social integration" (p. 4) when developing a theory on crisis tendencies. Social integration involves systems of institutions, such as school systems, integrating individuals into

society. The connection between social integration and system integration provides the basis for crisis tendencies.

Social integration responds to *social differentiation* in the lifeworld as students individuate with respect to the three worlds. For example, student progress in the cognitive-instrumental dimension leads to increased social differentiation in terms of division of labour and work in the economy (Durkheim, 1933/2012; Levy & Murnane, 2004). Student progress in the aesthetic-expressive dimension leads to differentiation in categories such as sexuality, gender and race (Foucault, 1984/1988, 1976/1990, 1984/1990; Gandhi, 1998; Irigaray, 1985, 1995; Said, 1978/1994; Spivak, 1985/2010; Wittig, 1993). Progress in the cognitive-instrumental and aesthetic-expressive dimensions is bound to progress in the moral-practical dimension through which social integration is facilitated and crises avoided. In this framing, Habermas (1976/1979) identifies crisis tendencies that extend beyond the economic and technical related crisis tendencies identified by Marx (1859/2000), to include crises in a social dimension that emerge from increased sociocultural differentiation in the lifeworld. Because of their direct relationship to education and educational assessment, crisis tendencies emerging from social differentiation and its relationship to social integration are of primary interest to this study's question on progress narratives.

Habermas (1981/1992, pp. 142-144) draws on Parsons (1951/2012) to conceptualise the lifeworld as involving the three reproductive processes of cultural reproduction, social integration and socialisation. In Habermas' (1981/1992, p. 140) framing, these processes engage with progress narratives and ensures that newly arising situations are integrated into existing conditions to secure continuity and coherence of daily life. Social integration addresses the orderly, cooperative, or conflictual relations between actors as concrete interactions in time and space (Mouzelis, 1997). Socialisation is the formative process that takes place in the medium "of grammatical language and from which both social institutions and the social identity of socialized organisms proceed with equal originality" (Habermas, 1981/1992, p. 24). Habermas (1981/1992) stresses the role of ordinary and everyday language for these reproductive processes, which is important in the context of education assessment in the lifeworld of the classroom.

System integration

System integration refers to the steering performance of self-regulating systems, “based on the systemic steering media of money and power that regulate actions more or less ‘automatically’” (Mouzelis, 1997, p. 114). Habermas (1992/1998) considers the function of law as anchoring system integration and social integration, where both the

media of systemic integration, money and power, are anchored via legal institutionalization in orders of the lifeworld, which is in turn socially integrated through communicative action. (p. 40)

This study explores educational assessment as a similar self-regulating symbolic media. In this way, educational assessment can be considered as similarly integrating with the economy through law when it is anchored in legally constituted institutions or in administrative processes of the state.

System integration involves coordination “based on the systemic steering media of money and power that regulate actions more or less ‘automatically’” (Mouzelis, 1997, p. 114). System level steering media provide the mode for system differentiation and integration and develops in modern societies where

increasingly autonomous organizations are connected with one another via delinguistified media of communication: these systemic mechanisms—for example, money—steer a social intercourse that has been largely disconnected from norms and values. (Habermas, 1981/1992, p. 154)

The proliferation of educational assessment use for systems coordination is a phenomenon associated in this study with progress narratives including the emergence of knowledge-based economies (Drucker, 1968/1992; Newton, 2007).

Crisis tendencies and legitimation

Habermas (1981/1992) conceptualises legitimation crises as emerging from tensions between social integration and system integration that respectively arise from lifeworld and system perspectives. This tension is associated here with system uses for educational assessment and uses in the lifeworld of the classroom. Legitimation crises involve crises

of meaning where systems coordinated through autonomous steering media clash with lifeworld processes that are coordinated through everyday language.

Habermas (1976/1979, p. 50) identifies four crisis tendencies for which he asserts completeness as a theorem. There are two system crisis tendencies, *economic crisis* and *rationality crisis*. There are also two identity crisis tendencies, *legitimation crisis* and *motivation crisis*. Habermas (1976/1979) provides preliminary explanations for each crisis tendency for which he doesn't claim completeness. Educational assessment pertains to identity crises and particularly legitimation crises which address the relationship between state sponsored activities such as educational assessment programs and citizens. Identity crises do not endanger system integration in the way that, for example, an economic crisis might (Habermas, 1975/2005). This is consistent with this study's identification of legitimation crises in Australian education that seem somewhat isolated from Australia's economy that has experienced twenty-five consecutive years of output growth (OECD, 2017a).

How identity crises emerge around schools is illustrated in Figure 5. Motivation crises emerge from social differentiation experienced in the lifeworld through division of labour arising from technological progress and through social differentiation around categories such as class, race, gender and sexuality. Social integration, in the context of this study, occurs through institutions such as schools where students, as well as their families and broader sociocultural structures, encounter each other and are integrated into society. Legitimation crises emerge when schools and other institutions in society are integrated through system generated symbolic media. Figure 5 illustrates how both identity crisis tendencies of motivation and legitimation emerge within and around schools.

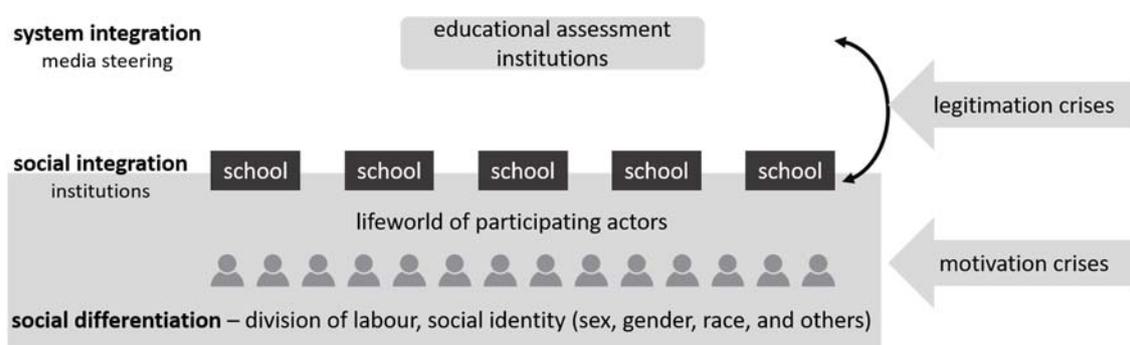


Figure 5 – How schools relate to motivation and legitimation crisis tendencies

Motivation crises are distinct from legitimation crises, and “are a result of changes in the socio-cultural system itself” (Habermas, 1975/2005, p. 48). Motivation crises may result from certain forms of social differentiation, or a failure of social integration through institutions such as schools which may fail to address the immediate structure of the cultural lifeworld. While this study does not focus on motivation crises, Chapter 6 contextualises a motivation crisis arising from differentiation on the category of sexuality and how this affects the legitimacy of social integration in schools (Rizmal, 2017; Rodwell, 2016).

Schools are implicated in motivational crises through concerns over how students engage with society and how students encounter the lifeworld inside and outside of school (Ames, 1992; Fredricks, Filsecker, & Lawson, 2016; Willms, 2003). Motivation crises also relate to psychopathologies such as ennui and anomie (Habermas, 1981/1992).

Legitimation crises pertain to the relationship that schools have with governments and central offices, a relationship that may be mediated by educational assessment data when used in school evaluation for example (Newton, 2007; OECD, 2013c). Legitimation tensions between schools and central offices also emerge when the steering media of money is combined in mixed ratios with the symbolic media of educational assessment through human capital theory (G. S. Becker, 1964/1993). Similar legitimation tensions emerge when money and educational assessment are combined to evaluate the productivity of education (Hanushek & Etema, 2017), and how education relates to national economic growth (Hanushek, 2013, 2016; Hanushek, Schwerdt, Wiederhold, & Wößmann, 2013).

Legitimation and legitimacy are addressed throughout this study in terms of how citizens relate to the administrative systems of the state. Legitimacy involves “the formally democratic securing of the private autonomous disposition of the means of production” (Habermas, 1975/2005, p. 46), where through social integration schools play a central role in fostering such a disposition. Legitimation crises emerge from tensions between system generated symbolic media and autonomous citizens in the lifeworld.

System steering in a knowledge economy

Legitimation crises in education emerge from education being increasingly identified as a driver of both citizen and societal prosperity in knowledge-based economies (G. S.

Becker, 1964/1993; Drucker, 1968/1992; Hanushek, 2013, 2016; Hanushek & Wößmann, 2010). As Chapter 4 identified, the proliferation of educational assessment use in recent decades includes what Newton (2007) describes as national accounting purposes, and is explored here through human capital theory (G. S. Becker, 1964/1993).

Human capital theory can lead to one type of legitimation crisis through the steering media of money having a different validity horizon to the symbolic media of educational assessment (G. S. Becker, 1964/1993). Habermas (1981/1992, pp. 122-123) describes lifeworld actors as being in situational contexts of moveable horizons. The horizons of educational assessment as perceived by lifeworld actors such as students have similarly moveable horizons. The potentially restricted horizon of educational assessment contrasts to the horizon of money which can theoretically be indefinitely extended “through anonymous market mechanisms” (Rehg, 1992/1998, p. xviii). A legitimation tension therefore emerges when human capital theory is applied to combine the two media through numerical ratios (G. S. Becker, 1964/1993; Hanushek, 2013, 2016; Hanushek & Wößmann, 2010).

The tension between system and lifeworld perspectives, as well as the potentially differing validity horizons of systems generated media and the medium of language used in the lifeworld, provides a useful framework for analysis. For example, the different validity horizons might explain the variability in the efficacy of human capital theory identified by Blaug (1976) and Tan (2014). Tan (2014), in a review of the literature on human capital theory, found “some empirical studies whose results are in discord with [human capital theory] and it is clear that [human capital theory] fails to explain those” (p. 436). This study does not empirically explore this tension other than to identify a legitimation tension between the media of money and educational assessment that might explain the variability in the efficacy of human capital theory observed by Tan (2014).

5.2. Educational assessment and symbolic media

Moss (2016) frames educational assessment validity through an intended interpretation that is reported and an actual interpretation in data use. This section explores educational assessment as symbolic media that mediates between these two interpretations. Moss’ (2016) framing has analogies in the field of sociology including the performance-sanction paradigm elaborated by Parsons and Smelser (1956/2005, p. 9) for intersystemic

communication. The performance-sanction paradigm is further explored in Chapter 6 and for the purposes of this study the performance-sanction paradigm is used interchangeably with intended and actual interpretation of educational assessment data.

There are several conceptualisations of symbolic media in the field of sociology with Parsons (1963a, 1963b), Luhmann (1992, 1984/1995) and Habermas (1981/1992) providing three distinct conceptualisations (Chernilo, 2002). While this study's point of departure is the conceptualisation by Habermas (1981/1992), the other conceptualisations provide useful insights.

Parsons (1963a) provides a useful point of departure in describing the medium of influence in terms of a symbol or category of symbols where actors accept "the 'symbolic' in lieu of the 'real'" (p. 42). In the context of educational assessment, the *symbolic* refers to media such as transcripts, reports and related media generated by educational assessment.

Habermas (1981/1992, pp. 258-259) identifies Parsons (1963a, 1963b) as conceptualising sixty-four types of media and argues that Parsons (1963a, 1963b) overgeneralises. Habermas (1981/1992) speculates that this overgeneralisation results from Parsons (1963a, 1963b) developing the concept of media on the "structural analogies to the money medium" (p. 258). Habermas (1981/1992, p. 154) considers money a special type of symbolic media that is disconnected from norms and values and used within increasingly complex networks without the need for anyone to comprehend or take responsibility for its meaning. Habermas (1981/1992) argues that

Media such as money and power attach to empirical ties; they encode a purposive-rational attitude toward calculable amounts of value and make it possible to exert generalized, strategic influence on the decisions of other participants. (p. 183)

Habermas (1981/1992) uses the term *steering media* for symbolic media such as money and administrative power which does not require reference to the lifeworld as "the meaning-frame for actors' action orientations" (Chernilo, 2002, p. 440).

Habermas (1981/1992, pp. 390, 277) distinguishes steering media from *generalised forms of communication*, or *generalised communication*, which requires recourse to the

lifeworld for meaning. Generalised communication simplifies and condenses overly complex relationships between communication and action and is distinct from money which does not require recourse to linguistic argument to establish meaning. Habermas (1981/1992, pp. 280-290) considers media such as influence and value commitment, generalised communication for which validity and meaning are redeemed through ordinary language referencing a lifeworld context.

Habermas (1981/1992, pp. 280-290) emphasises that steering media is distinct from generalised communication. When symbolic media is considered as steering media such as money, then actors using it in exchange need to take no responsibility for its meaning. When symbolic media is considered as generalised communication, meaning cannot be decoupled from the lifeworld context in which it is used. However, Habermas (1981/1992) also describes how the demands on everyday language can become overloaded to be “replaced by delinguistified media” (p. 155). Habermas (1981/1992) associates this development with a “tendency toward an uncoupling of system and lifeworld” (p. 155). Whether educational assessment is considered as steering media removed from the lifeworld, or a form of generalised communication where meaning is attached to a lifeworld, pertains to the study question on how educational assessment is used in society.

Institutional and lifeworld anchoring of symbolic media

Parsons (1963a, 1963b) and Habermas (1981/1992) each takes a distinct approach to how the symbolic media of influence is anchored. Each approach is manifest in the literature on educational assessment and both are addressed here.

Parsons (1963a, pp. 41-42) describes influence as institutionally anchored, which is a conceptualisation challenged by Habermas (1981/1992). Nevertheless, educational assessment is often anchored in institutions such as the OECD (OECD, 2000). Secondary school exit credentials are also often institutionally anchored, such as those administered by the Victorian Curriculum and Assessment Authority (VCAA) (www.vcaa.vic.edu.au). Parsons (1963a) describes institutional anchoring of symbolic media as requiring: a *category of value* over which there is something at stake for actors; a *category of interest* where the category of value applies to action; a *definition of the situation* in which the category of interest can be exploited; and a *normative framework of rules* for

discriminating between legitimate and illegitimate exploitations of the media. In this way, Parsons (1963a) provides a rudimentary framing for educational assessment as institutionally anchored symbolic media.

Habermas (1981/1992, p. 275) challenges Parsons (1963a) to argue that the medium of influence is not sufficiently well circumscribed to permit it being anchored in institutions. Instead, Habermas (1981/1992) argues that influence is instead anchored in the normative context of the lifeworld. There are educational assessment practices that are anchored in the lifeworld, particularly school-based and classroom-based assessments that address how groups of students and their teachers relate to educational content (Cowie & Bell, 1999; Harlen & James, 1997; Moss, 2003; Shepard, 2006). Habermas' (1981/1992) contrasting position to Parsons (1963a, 1963b) provides an additional useful framing for educational assessment, and both are used throughout this study.

The forms of anchoring also pertain to differences in evidence models described in Chapter 2. Classical test theory is anchored in a lifeworld and corresponds more closely to how symbolic media of influence is conceptualised by Habermas (1981/1992). Item response theory is more amenable to institutional anchoring similar to how Parsons (1963a) describes symbolic media. In this way, the selection of mathematical model relates to how the meaning of educational assessment is anchored.

Educational assessment as symbolic media

The concept of educational assessment as symbolic media provides a useful conceptual tool for progressing the research agenda of Moss (2016). Moss (2016) considers intended interpretation the conventional focus of validity (Kane, 2006; Messick, 1989). The second is actual interpretation and is made by teachers, administrators, policy-makers and others leading to consequences and actions. Moss (2016) argues for an increased focus on actual interpretation.

Shepard (2016) makes a similar observation to Moss (2016), that actual interpretation can be quite removed from intended interpretation for which different approaches to validation are required. Shepard (2016) suggests that educational assessment and technically derived terms such as reliable, valid and fair are developed by experts and used

as a short-hand to communicate with other professionals, courts and judges, parents and other test users. (p. 271)

Shepard (2016) argues that for test users such as courts and policy-makers the question of validity centres on whether “a test is adequate for its purpose” (p. 271).

Shepard’s (2016) argument bears on this study in three ways. First, the argument is consistent with the notion that educational assessment is a form of generalised communication and a “short-hand” way for experts to communicate with other spheres of society on complex matters. Second, it is consistent with the argument that there are generally two interpretations of student responses in educational assessment, one when the educational assessment media is generated, the other when the media is used for decision-making such as by courts. Third and finally, Shepard (2016) illustrates how actual interpretation can occur outside of the sphere of education to involve fields of expertise outside of education.

5.3. Colonisation and separation of system and lifeworld

Habermas (1981/1992) describes the lifeworld perspective as one of an actor as an active interpretive participant in society, and the system perspective as that of an external observer who perceives society as a functional system. The interpretive participant is described by Habermas (1981/1992, pp. 123-124) as being in the centre of a lifeworld that is always present. This lifeworld has movable horizons and forms the background for the interpretive participant’s meaning-making where meaning is generated through the medium of language. In contrast, the symbolic media generated by systems have distinct characteristics and horizons and are autonomous so that they can become decoupled from ordinary language used in the lifeworld. Habermas (1981/1992) refers to the tendency of system generated steering media becoming uncoupled from ordinary language used in the lifeworld as *colonisation of the lifeworld*.

Colonisation describes the process where system imperatives and system generated media begin to impinge on normal reproductive processes of the lifeworld (Habermas, 1981/1992). In the context of educational assessment system generated media refers to that generated by programs such as the PISA and NAPLAN, and colonisation refers to the process where such media impinges on classroom processes. The literature review in

Chapter 2 identified several clear cases that illustrate colonisation resulting from system generated educational assessment. The study by Solomon and Lewin (2016), as well as the study by Hardy and Lewis (2016), identify teachers needing to address contradictory logics of system imperatives and imperatives emerging from the lifeworld of the school. Shepard (2000) identifies high-stake system level accountability testing as shaping and distorting instructional activities. These studies provide clear cases to support the hypothesis that system imperatives communicated through educational assessment can impose on activities in the lifeworld of schools.

The uncoupling of system and issues of legitimacy

Habermas (1981/1992) characterises the uncoupling of system and lifeworld emerging from increasing complexity generated by progress narratives. Increasing societal complexity leads to a proliferation of systemic mechanisms that become detached from social structures. Habermas (1981/1992, pp. 155-175) theorises the development of societies from tribal through to modern ones. For tribal societies, lifeworld structures are linguistically mediated and normatively guided through social structures. For example, in tribal societies positions of leadership might be claimed by older members of the group, or through status systems based on prestige. As societal complexity increases, kinship groups begin to segment and stratify, and steering media and state organisations begin to develop. The modern idea of citizen – bounded by laws and regulations – emerges from this elaboration (Habermas, 1981/1992, p. 170). As societal complexity increases, the tendency for system imperatives to become differentiated from the lifeworld experienced by citizens also increases.

Habermas (1981/1992, pp. 174-175) describes advanced societies as developing complex institutional structures. In tribal societies kinship structures integrate members into the clan. In advanced societies rights are maintained and justice is administered through institutional structures of increasing complexity. These institutional structures begin to embody higher levels of moral development to focus on coordinating society rather than being superordinate to it. As societies advance, Habermas (1981/1992) observes a “trend towards the growing dependence of legitimate orders on formal procedures for positing and justifying norms” (p. 174). This study addresses the legitimacy of the system generated educational assessment data that emerges from progress narratives.

Habermas (1981/1992) describes symbolic media as emerging when “ever greater demands are made upon this basic medium of everyday language” (Habermas, 1981/1992, p. 155). When everyday language becomes overloaded, it can be condensed using generalised communication or replaced by steering media. Steering media allows understandings that are recurrently established through communicative action to be replaced by social norms. Recurring understanding can also be replaced by steering media, and where appropriate integrated institutionally through functions of law. Steering media enables interactions between actors in space and time to become increasingly complex and autonomous (Habermas, 1981/1992, p. 184). The steering media generated by systems therefore have broader horizons and are distinct from the everyday language used for coordinating the lifeworld.

Habermas (1981/1992) characterises steering media as becoming increasingly abstract so that actors do not need to take responsibility for its meaning. In this way, steering media is distinguished from generalised communication whose validity needs to be linguistically redeemed when applied in concrete situations. This distinction between abstract steering media and generalised communication is further explored in Chapter 6 in relation to educational assessment.

Colonisation of the lifeworld

System generated steering media rationalises the lifeworld and uncouples the lifeworld from systems. As system generated steering media proliferates

the irresistible irony of the world-historical process of enlightenment becomes evident: the rationalization of the lifeworld makes possible a heightening of systemic complexity, which becomes so hypertrophied that it unleashes system imperatives that burst the capacity of the lifeworld they instrumentalize. (Habermas, 1981/1992, p. 155)

Habermas (1981/1992, p. 196) describes colonisation of the lifeworld arising when steering media disrupts symbolic reproduction of the lifeworld. Using Weber’s (1915/1964, 1920/2011) notion of rationalisation of society as a point of departure, Habermas (1981/1992) describes colonisation of the lifeworld as generated “by system imperatives that drive moral-practical elements out of private and political public spheres

of life” (p. 325). In this way, colonisation is considered “as a reification of the communicative practice of everyday life” (p. 386).

For Habermas (1981/1992), “media-steered subsystems develop *irresistible inner dynamics* that *bring about* both the colonization of the lifeworld and its segmentation from science, morality, and art” (p. 331). Habermas (1981/1992, p. 330) does not associate segmentation with differentiation according to value spheres, and instead associates it with expert cultures that split-off from the communicative context of daily life. Of interest to this study are the expert cultures and inner logic of educational assessment and its relationship with colonisation. This dynamic is particularly pertinent to the Australian context which has an expert culture in the Rasch (1960/1980) model as described in Chapter 1 (Adams et al., 1997; Andrich, 1988; Masters, 1982). In this sense, the principle proposed by Masters (2013) focusing on “where learners are” (p. 6) could be characterised as a manifestation of the inner logic associated with expertise around the Rasch (1960/1980) model and its ruler metaphor (Wright, 1997).

Habermas (1981/1992, p. 141) identifies a range of psychopathologies resulting from colonisation of the lifeworld and addresses them through Parsons’ (1951/2012) “domains of culture, society, and personality” (Habermas, 1981/1992, p. 142). Disturbances in cultural reproduction result in a loss of meaning and crises in the life orientation of individuals. Disturbances in social integration across society can result in unsettled collective cultural identities, anomie, and alienation. Disturbances in socialisation rupture an individual’s cultural tradition that may lead to further psychopathologies, particularly those associated with the ego. These pathologies are of interest to educators, as education and schools provide an opportunity to mitigate adverse effects arising from system imperatives.

The association that Habermas (1981/1992) makes between colonisation and psychopathologies is explored here through the empirical work of Howell (2012, 2017) among others. However, the empirical evidence produced by Bowman et al. (2017) as well as Rickwood et al. (2014), suggest a broader contemporary mental health crisis among Australian youth (McGorry et al., 2013). While educational assessment may also be implicated in this broader crisis, this relationship is not pursued here, and instead this study restricts its analysis to the study by Howell (2012, 2017).

The concept of communicative action and the active participant in the lifeworld is addressed next.

5.4. The active participant in the lifeworld

The use of system generated symbolic media for intersystemic coordination contrasts to how symbolic media is used in the lifeworld of the classroom. Habermas (1981/1985) builds on earlier theories of action developed by Weber (1915/1964) to develop pragmatic approach communication in the lifeworld. Weber (1915/1964, p. 115) identified four types of social action based on modes of orientation. First is action rationally oriented towards a goal that involves either external objects or other rationally oriented humans. Second is action rationally oriented towards a belief system drawn from ethical, aesthetic, religious or other system of beliefs. This form of action is taken for its own sake without reference to a criterion of success or a goal. Third is action motivated by an affectual orientation such as an emotion or feeling. Fourth, and finally, is habituated action oriented towards a tradition. For Habermas (1981/1985), Weber's (1915/1964) four forms of action somewhat presuppose a cultural context and common understanding, and Habermas (1981/1985) considers this formulation inadequate in developing a pragmatic approach.

Habermas (1981/1985, p. 281) considers Weber's (1915/1964) framing too subjectively based. It presumes common cultural understandings upon which strategic and goal-oriented action between individuals are based. That is, Weber (1915/1964) presumes that two goal-oriented actors understand each other to allow for rational goal-oriented action. Habermas (1981/1985) finds this presumption does not hold in contemporary contexts and responds by developing the concept of communicative action as a way of establishing common background understanding.

Habermas' (1981/1985) communicative action addresses how actors understanding each other's rationality before engaging in strategic goal-oriented action. In conventional societies, this background understanding is provided through a shared culture. In pluralistic societies, these background understandings first need to be linguistically established before strategic action can be embarked upon. For Habermas (1981/1985, 1981/1992), the formation of common background understanding is achieved through communicative action from a lifeworld perspective using ordinary language.

Habermas (1981/1985, p. 285) reworks Weber's (1915/1964) subject-centred theory of action into a theory with three basic types of action. First, *instrumental action* is action oriented towards success in non-social situations. Second, *strategic action* is action that is goal-oriented towards success in social contexts. Third, *communicative action* is action oriented towards generating common understanding in social contexts. Habermas' (1981/1985) instrumental action and strategic action describe the goal-oriented actions of Weber (1915/1964). Habermas' (1981/1985) essential contribution is communicative action that provides for the formation of intersubjective understanding presupposed in Weber's (1915/1964) framing.

Habermas (1981/1985) considers instrumental action and strategic action to be parasitic or dependent on communicative action. This assertion is based on the argument that actors need to reach a common understanding of each other's rationality before being able to strategically act towards each other (Habermas, 1981/1985, p. 288). The notion that other modes of language are parasitic on communicative action is contested however (Rorty, 1995, p. 443; Saurette, 2005, p. 178). Luhmann (1992), for example, argues that systems can continue to survive and develop without recourse to valid forms of communication. Further, Cooke (1997) argues that "there is no evidence that Habermas has adequately addressed this issue" (p. 24) to also contest the primacy of communicative action in society. Nevertheless, Cooke (1997) goes on to observe that the primacy of communicative action can be defended at the conceptual level, while not the functional one. Further, Cooke (1997) considers that a choice between strategic or communicative action may be available to individual actors engaged in exchange, but that this choice does not exist for creating common background understanding. Cooke's (1997) point on the need for communicative action for generating background understanding pertains to this study's question over how new common understandings are generated within the broader progress narratives in society.

The concept of communicative action is central to this study's approach where the conceptualisation by Habermas (1981/1985) provides a coherent link to other important issues this study identifies as affecting educational assessment design. In particular, the work of Austin (1962/1975) provides a bridge between communicative action (Habermas, 1981/1985) and the concept of performativity (Ball, 2003; Lyotard, 1979/1984) and to the construction of identity (Butler, 1988, 1990/2007). These are explored next.

Words as action – communicative action and performativity

Habermas (1981/1985) draws on Austin's (1962/1975) *How to do Things with Words* to elaborate how different forms of speech relate to communicative action and strategic action. Austin's (1962/1975) theory of speech is developed around three speech acts that emerge from speech. First is the act of *locution* that communicates something, it is the "act of 'saying something'" (p. 94). How a locution or message is understood depends on how the locution is delivered and received which pertains to two other acts of speech.

The *illocutionary* act is the second speech act and is a "performance of an act *in* saying something as opposed to performance of an act *of* saying something" (p. 99). In this sense, a speech act can be made with *illocutionary force* that is a function of the circumstances of an utterance. For example, increased illocutionary force gives rise to verdictives and exercitives that are performatives with a relationship to law and authority. Illocutionary force and the way something is said may also, for example, render an illocution a demand rather than a request.

The third speech act is the *perlocutionary act* that has a consequence for the hearer and audience. Saying something generally has certain consequential effects upon the feelings and thoughts of the audience, and even perhaps the speaker. This effect can be achieved through design and intention, or the effect may be unintentional. Further, the perlocutionary effect of a speech act might only be obliquely related to the locution or the content of the speech act itself (Austin, 1962/1975, p. 101).

Austin (1962/1975) makes clear that the effect of saying something may or may not relate to the propositional content of what is said. In the context of educational assessment, this means that the perlocutionary effect of an actual interpretation may not necessarily resemble the propositional content of the intended interpretation. For example, maladaptive perlocutionary effects from NAPLAN were found in the empirical study by Howell (2017). These maladaptive responses might not directly emerge from the intended interpretation of NAPLAN in terms of propositional content. Austin's (1962/1975) work suggests that these maladaptive responses may be due to either the illocutionary force of reporting, or the psychological disposition of the child.

Austin (1962/1975, pp. 5-7) develops the concept of the performative utterance, or simply the performative, to describe how certain speech acts not only describe but perform to

have an effect. Austin (1962/1975) distinguishes “performative utterances from constative” (p. 94), where performative speech that creates is distinguished from speech that simply describes or constates. Butler (1988, 1990/2007) uses the concept of the performative to show how speech can constitute gender. For Lyotard (1979/1984), the distinctive feature of the performative utterance “is that its effect upon the referent coincides with its enunciation” (p. 9). Habermas (1981/1985) uses the notion of the performative to distinguish various forms of social action.

Habermas (1981/1985) uses Austin’s (1962/1975) notions of locution, illocution and perlocution to distinguish between communicative action and strategic action. For Habermas (1981/1985), the *communicative act* is where the

self-sufficiency of the speech act is to be understood in the sense that the communicative intent of the speaker and the illocutionary aim he is pursuing follow from the manifest meaning of what is said. (p. 289)

For Habermas (1981/1985) therefore, for an act to be communicative it must have appropriate illocutionary force. The communicative act drifts into strategic action when

perlocutionary effects which result from the fact that illocutionary acts sometimes take on roles in contexts of strategic interaction. These effects ensue whenever a speaker acts with an orientation to success and thereby instrumentalizes speech acts for purposes that are only contingently related to the meaning of what is said. (p. 289)

Habermas (1981/1985) further distinguishes between open strategic and concealed strategic actions. Concealed strategic actions are

interactions in which at least one of the participants is acting strategically, while he deceives other participants regarding the fact that he is not satisfying the presuppositions under which illocutionary aims can normally be achieved. (p. 294)

A concealed strategic act may be unconscious on behalf of both hearer and speaker, leading to *systematically distorted communication*. Habermas (1981/1985) considers this a pathology that “can be conceived of as the result of a confusion between actions oriented to reaching understanding and actions oriented to success” (p. 332). In contrast,

manipulation is at a conscious level and occurs when “at least one of the parties behaves with an orientation to success but leaves others to believe that all the presuppositions of communicative action are satisfied” (p. 332). These strategic goal-oriented forms of action are all distinct from communicative action.

Figure 6 illustrates the framework of social action that emerges from Habermas’ (1981/1985) reformulation of Weber’s (1915/1964) approach to action. The framework illustrates the distinction between instrumental and social action in educational assessment design and use, with this study addressing the question on how educational assessment *ought* to be designed as a form of social action.

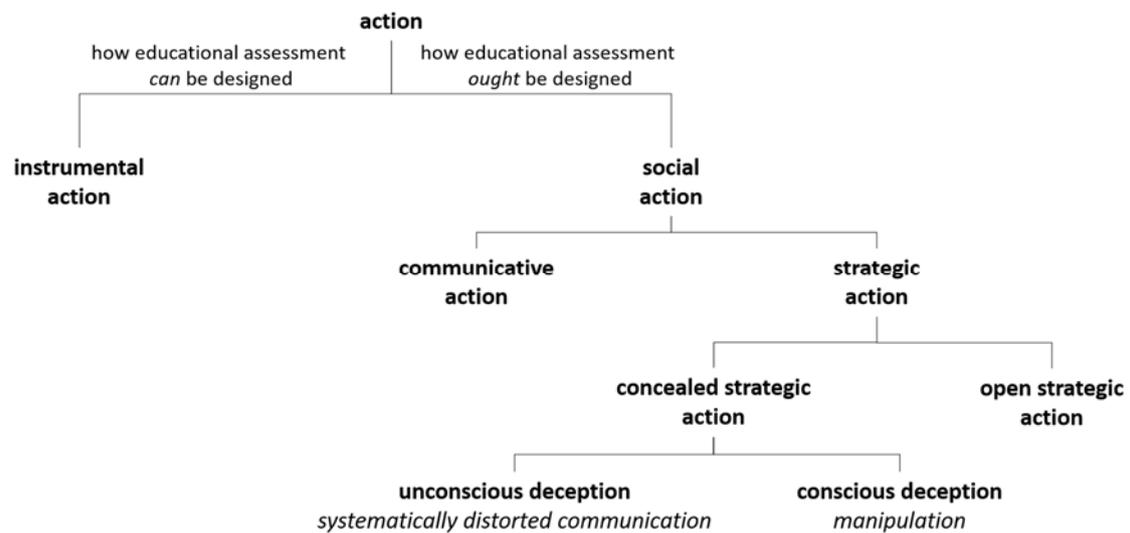


Figure 6 – Framework of social action used in this study
Habermas (1981/1985)

Social action and educational assessment

The conceptualisation of social action developed by Habermas (1981/1985) can be mapped onto practices in educational assessment. For example, the framing of intended interpretation and actual interpretation provided by Moss (2016) can be mapped onto elements of speech acts described by Austin (1962/1975). The intended interpretation can be considered analogous to the locution or the message that is sought to be communicated by reporting, and the perlocutionary effect arising in the audience from that message can be considered analogous to the actual interpretation. In this conceptualisation, how an intended interpretation is reported is a function of both the locution and the illocutionary

force of the speech act. In this way, Habermas (1981/1985) provides a useful way of exploring how the reporting of educational assessment can be either communicative or strategic.

Teaching practices include both communicative and open strategic action. That education involves strategic action that seeks to have an effect on students, provides a more formal argument that what is assessed by educational assessment is intersubjectively constructed. When acting communicatively, a teacher is seeking to make something known. When acting strategically, a teacher might purposefully seek to elicit a certain response in the student, such as a learning response. Strategic action that results in a learning response can be cast into Austin's (1962/1975) framing as a perlocutionary effect. Teaching becomes openly strategic when the teacher informs the student of the effect the teacher is seeking to have. In this sense, teaching becomes openly strategic when intentions are openly declared through curriculum or syllabus and when a teacher's actions are consistent with these statements. In the absence of such a declaration, a teacher's action could be considered concealed strategic action where communication might become either distorted or manipulative.

Educational assessment as social action

Educational assessment provides one form of strategic communication between the teacher and the student. Engelmann, Becker, Carnine, and Gersten (1988), for example, describe a direct instruction model for education that includes scripted reinforcement for correct responses. Engelmann et al. (1988) provide examples of reinforcement such as praise and suggest not to "use a stronger reinforcement system than necessary to get the job done" (p. 306). In the context of this study, Engelmann et al. (1988) are bringing to attention the illocutionary force of praise, and that if given with inappropriate force the message intended to be conveyed by praise might become distorted. Engelmann et al. (1988) make a similar argument for correcting student mistakes, and to minimise the illocutionary force by communicating to simply "remind a student of the process to follow to determine the correct answer" (p. 306). Here, "remind" could be contrasted to, for example, "chastise" in terms of illocutionary force and possible perlocutionary effects. These examples illustrate the importance of illocutionary force in the context of

educational assessment and how action can drift into distorted communication or manipulation.

Educational assessment and the construction of identity

The concept of the performative utterance and its perlocutionary effect foregrounds the capacity of educational assessment to create and construct something in students. Lipnevich et al. (2016) describe feedback as an involved process that can have either an adaptive or a maladaptive effect in students. Howell's (2017) study also suggests that reporting can lead to systematically distorted communication. The studies by Lipnevich et al. (2016) and Howell (2017) illustrate how educational assessment can be used to construct student identity.

Butler (1988; 1990/2007, p. 136) addresses the construction of gender identity through acts, gestures, and other performative enactments. Butler's (1988, 1990/2007) conceptualisation can be extended to the construction of identity along cognitive, social and aesthetic dimensions through educational assessment. Chapter 6 further explores how educational assessment is implicated in the construction of identity. Butler (1988, 1990/2007) alerts to possibility that the reporting of an intended interpretation can create, reinforce, or distort identity. Further, both Butler (1988, 1990/2007) and Austin (1962/1975) show how construction is dependent on the intended interpretation, how the intended interpretation is reported through its illocutionary force, and the context in which the actual interpretation is made.

Unobtrusive data collection

The theory of communicative action developed by Habermas (1981/1985) also provides a framing for ubiquitous and unobtrusive data collection described by Behrens and DiCerbo (2014). Mainly, if the unobtrusive nature of data collection conceals that data collection from the student, communication can become systematically distorted or manipulative. This makes unobtrusive data collection problematic in education as well as other spheres such as marketing (Achrol & Kotler, 2011). This is further addressed in Chapter 6.

Communicative action and strategic action are not only relevant to the reporting of intended interpretation, but also important for creating common understanding in

educational assessment design and implementation. The use of communicative action in general discourse is addressed next.

5.5. Discourse ethics

Communicative action is central to Habermas' (1981/1985, 1981/1992, 1992/1998) program of universal pragmatics. For Habermas (1981/1992), communicative action

relies on a cooperative process of interpretation in which participants relate simultaneously to something in the objective, the social, and the subjective worlds, even when they *thematically stress only one* of the three components in their utterances. (p. 120)

Habermas' (1981/1985, 1983/1996, 1998) pragmatics, communicative action, and discourse ethics therefore presupposes the pattern of three worlds, with all three worlds implicated in communication even only one is stressed.

Habermas (1983/1996) draws on Toulmin's (1950/1970, 1958/2003) approach to argument to determine how validity is established in communication. As described in Chapter 3, Toulmin (1958/2003, p. 116) distinguishes between analytic and substantial arguments. For analytic arguments, "the backing for the warrant authorising it includes, explicitly or implicitly, the information conveyed in the conclusion itself" (Toulmin, 1958/2003, p. 116). For example, the argument "Anne is one of Jack's sisters; All Jack's sisters have red hair; So, Anne has red hair" (Toulmin, 1958/2003, p. 115) is an analytic argument. In this form of argument all the backing for the argument is contained within the argument. Once it is accepted that Jack's sisters have red hair, and that Anne is one of Jack's sisters, it is logical to conclude that Anne has red hair. This argument is therefore considered somewhat tautological.

Toulmin (1958/2003, p. 116) considers an argument substantial when the argument is not tautological, and when data provides the basis for a claim, conclusion or predicate. For Toulmin (1958/2003), substantial argument involves claims that require supporting warrants and claims that remain open to rebuttal through further evidence. This may lead to a claim or conclusion being qualified. Toulmin's (1958/2003) approach is illustrated in Figure 7.

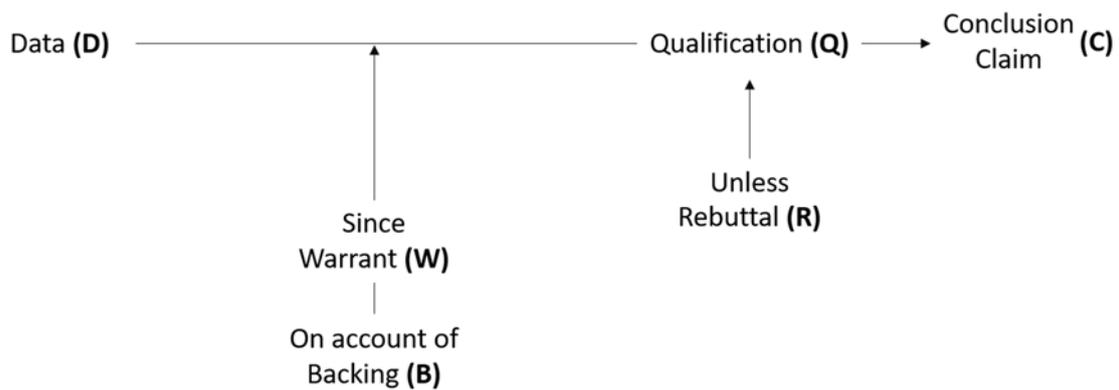


Figure 7 – The Toulmin approach to arguments
Toulmin (1958/2003)

Habermas (1983/1996, pp. 56-57) is attracted to Toulmin’s (1958/2003) approach because it abandons semantic analysis of expressions and sentences, and instead focuses on a mode through which normative propositions are justified. Toulmin (1950/1970) addresses the question on what “kinds of argument, what kinds of reasoning is it proper for us to accept in support of moral decisions” (p. 67). The Toulmin (1958/2003, p. 15) approach to argument also accommodates a plurality of meaning and modes that arise from arguments in different fields to accommodate, for example, legal arguments in courts and other kinds of argument in say education or sport. Here, Toulmin (1958/2003, p. 15) develops a distinction between *field-invariant* and *field-dependent* forms of argumentation that allow scientific and ethical warrants and rebuttals to be simultaneously considered. Toulmin’s (1950/1970, p. 74; 1958/2003, p. 63) approach addresses what is beauty, what is good, what is right, what is valued, what is valid, and what is worthy of belief. Habermas (1983/1996, p. 53) argues that these predicates are of a higher-level than claims addressed by analytic arguments.

The difference between analytical and substantial arguments is significant for evidence-based decision-making and policy development that use symbolic media generated by educational assessment. Gorur and Wu (2015), for example, question the development of public policy using country rankings reported by the PISA. Gorur and Wu (2015) go on to explore the content of the PISA test to challenge the validity of rank-based predicates to policy arguments. That is, it’s a critique against the analytical logic of “rank X is better than rank Y, we are rank Y, therefore we must aim for rank X”. In terms of Toulmin (1958/2003), Gorur and Wu (2015) are seeking to shift the policy argument from an

analytic argument based on the category of ranks, to a substantive argument about what is good, right and true with respect to test content. Gorur and Wu (2015) also address the connection between the linguistic and numerical components of educational assessment, by raising substantial arguments about linguistic meaning of educational content, rather than analytical argument about numerical ranks.

The difference between substantial and analytic arguments directly relates to performativity. Lyotard (1979/1984) argues that

Data banks are the Encyclopedia of tomorrow. They transcend the capacity of each of their users. They are "nature" for postmodern man. (p. 51)

This sentiment can be cast in terms of an excessive focus on analytic arguments pertaining to the analysis of databases such as those from the PISA (OECD, 2016b). In terms of the postmodern condition, this focus on analytic arguments comes at the expense of engagement with substantial arguments addressing justice, beauty and truth (Lyotard, 1979/1984, p. 19), to lead to performativity.

Analogous claims to validity

Habermas (1983/1996, p. 56) argues that normative claims to validity are analogous to scientific truth claims; with the former requiring moral justification and the latter theoretical explanation. Habermas (1983/1996, p. 51) considers moral arguments as drawing on a web of moral feelings and attitudes, and that these arguments are analogous to scientific arguments that draw on a stream of perceptions. Further, moral criticism serves to correct judgements about modes of action, in a similar way to how theoretical criticism serves to correct beliefs about misleading experiences and perceptions. Here too, Habermas (1983/1996, pp. 51-57) draws on Toulmin (1950/1970, pp. 122-124), who argues that questions of physical and moral reality have parallel forms of argument, even if the modes and types of arguments presented are field-dependent. Questions over physical reality and moral reality respectively begin with reports of sensations and feelings that, once reported, become open to challenge on various grounds through communicative action. In this way, sensations and feelings provide discursively redeemable claims to validity in the disciplines of science and ethics respectively (Habermas, 1983/1996, p. 59).

Validity and communicative action are central to discourse ethics. Habermas (1981/1985) describes the process of establishing common understanding between actors as involving

(a) the three world-relations of actors and the corresponding concepts of the objective, social, and subjective worlds; (b) the validity claims of propositional truth, normative rightness, and sincerity or authenticity; (c) the concept of a rationally motivated agreement, that is, one based on the intersubjective recognition of criticizable validity claims; and (d) the concept of reaching understanding as the cooperative negotiation of common definitions of the situation. (p. 137)

In argument, how well an utterance fulfils its representational function of the objective world is measured against truth conditions. Habermas (1988/1998) argues that for the social and aesthetic worlds “the fulfilment of the expressive and the interactive functions is also measured against conditions that are *analogous to truth*” (p. 295). For Habermas (1981/1992, p. 120), these worlds are simultaneously accessed in speech from the totality of the subjective world of the speaker to provide the basis for communicative action.

The universal in discourse ethics

Habermas (1981/1985, pp. 34-35) identifies limitations in Toulmin’s (1958/2003) approach that pertains directly to this study’s claim that the traditional treatment of validation in educational assessment is similarly limited. While Toulmin (1958/2003) provides the logic for argumentation, he doesn’t conceptualise a standpoint of impartiality and procedures for a cooperative search for normative truth. Toulmin (1958/2003) recognises that validity claims are established by community-produced consensus, but Habermas (1981/1985, p. 35) considers that this doesn’t sufficiently differentiate between warranted and unwarranted consensual decisions. This criticism can also be applied to the approaches towards validation by Kane (2006) and Mislevy, Steinberg, and Almond (2003), who both use Toulmin (1958/2003). Habermas (1983/1996) develops an ethical approach that orients discourse towards the universal to address this limitation in Toulmin (1958/2003).

Habermas (1983/1996, pp. 76-109) uses a hypothetical argument with a sceptic to extend Toulmin’s (1958/2003) approach. The essence of the argument is that “Habermas wants to show that, if a person participates in communication, that person presupposes the

validity of not just some set of norms, but a particular set of universal norms” (Braaten, 1991, p. 43). Habermas’ (1983/1996) imagined victory over the sceptic largely rests on the notion of a *performative contradiction* that occurs when the propositional content of an argument contradicts the presuppositions asserting it. For example, a declaration by a person that they are not participating in an argument is a performative contradiction. Here, Habermas (1983/1996) draws on Watt (1975), who argues that to

raise a question is to be logically committed to its presuppositions, direct and indirect. To attempt to raise a question while repudiating any of its presuppositions is to talk nonsense. (Watt, 1975, p. 41)

There is a sense Habermas (1983/1996) is addressing those he considers postmodern (Habermas, 1981, 1997) in developing the notion of a performative contradiction. This sense particularly emerges from Lyotard (1979/1984, p. 60), an interlocutor of Habermas (1983/1996), who considers consensus an inadequate concept and who argues in favour of paralogical reasoning (Rorty, 1984).

Saurette (2005) makes a justifiable challenge in asserting that Habermas’ engagement with the sceptic is not in “the language of engagement or dialogue” (p. 177). The imagined debate is artificial. Habermas (1983/1996) himself recognises that the process of “reaching a rationally motivated agreement, must satisfy improbable conditions” (p. 88). It is upon these improbable conditions that Habermas (1983/1996) attempts to describe the “characteristics of an ideal speech situation” (p. 88). From a theoretical perspective, Habermas’ (1983/1996) argument is feasible, but can be difficult to envisage in concrete situations. Habermas (1983/1996) goes on to argue that the ideal speech situations require the development of postconventional reasoning addressed next.

Postconventional reasoning and ideal role taking

To develop the concept of ideal speech and ideal role taking, Habermas (1983/1996) uses Kohlberg’s (1971) model of moral development. It is through Kohlberg (1971) that Habermas (1976/1979) conceptualises development in the moral-practical dimension through which this study responds to the imperative by Elwood (2013) for greater attention to ethics in educational assessment design and use.

Kohlberg (1971) describes modern societies as having postconventionally structured domains of action, universal doctrines of legitimation, and approaches to resolving conflict that entail a strict separation of legality and morality. Kohlberg (1971) defines the postconventional level of moral development as a stage where

there is a clear effort to define moral values and principles that have validity and application apart from the authority of the groups or persons holding these principles and apart from the individual's own identification with these groups. (p. 77)

In this sense, postconventional reasoning provides for a decentred understanding of the world through discourse. That is, a perspective where understanding is not centred on the orientation of a single actor.

Habermas (1983/1996, p. 118) develops the concept of ideal role taking to address plurality in terms of cultures that may have different conceptions of morality. The notion of ideal role taking is important, for example, for participation in forums in educational assessment design, such as domain expert groups in the PISA (OECD, 2000, p. 103). These participants in international forums need to adopt a decentred ideal role when the outcomes of argument affect actors in various nations, countries and cultures. Cross-cultural coordination evident in the PISA requires recourse to ideal speech and ideal role-taking.

Following Kohlberg (1971), Habermas (1983/1996, p. 182) develops the notion of ideal role taking to signify a procedural type of justification. Ideal role taking becomes possible at the postconventional and principled stage of development. Habermas (1983/1996) uses the notion of ideal role taking to describe two principles that characterise sound moral judgement. For Habermas (1983/1996), the principle of universalisation (U), and the principle of discourse ethics (D), ensure that agreed norms are acceptable to all affected regardless of culture or background if

(U) All affected can accept the consequences and the side effects its general observance can be anticipated to have for the satisfaction of everyone's interests (and these consequences are preferred to those of known alternative possibilities for regulation). (p. 65)

...

(D) Only those norms can claim to be valid that meet (or could meet) with the approval of all affected in their capacity *as participants in a practical discourse*. (p. 66)

It is through these two principles that Habermas (1983/1996) extends Toulmin's (1958/2003) approach to anchor normative validity claims from a standpoint of impartiality. These two principles provide Habermas' (1983/1996, p. 32) weak transcendental approach that contrasts to the transcendental unity proposed by Kant (1781/2016). Habermas (1994, 1992/1998) goes on to extend these principles for application through deliberative democracy that involve processes of justification and application. This extension can also be applied to evidentiary reasoning assessment (Mislevy, Steinberg, & Almond, 2003) and validation in educational assessment (Kane, 2006). Once this extension is made, the concept of educational assessment validity begins to incorporate notions of legitimacy to allow educational assessment design and use to better respond to the needs of society and its citizens.

Habermas (1983/1996, pp. 187-188) provides a cognitive approach to argument and discourse and concludes his discussion by raising the possibility of psychological inhibiting factors. These are not further explored by Habermas (1983/1996) however these factors might be important for educational assessment design and use. While psychologically inhibiting factors are not explicitly addressed by this study, a review of the literature identified *object relations theory*, and its associated concepts of splitting, projection and identification, as a potentially useful framework (Bion, 1963, 1961/1996; Fairbairn, 1952/1994; Hinshelwood, 1987; Rioch, 1970). The psychological concepts of splitting, projection and identification provide an alternative explanation for the unhelpful dichotomies identified by Masters (2013), and therefore an alternative explanation for the malaise in educational assessment.

5.6. Deliberative democracy

Habermas (1992/1998) seeks to ameliorate the effects of colonisation of the lifeworld through deliberative democracy which institutionalises procedures promoting

communicative action (Habermas, 1981/1985) and discourse ethics (Habermas, 1983/1996). For Habermas (1992/1998)

the success of deliberative politics depends not on a collectively acting citizenry but on the institutionalization of the corresponding procedures and conditions of communication, as well as on the interplay of institutionalized deliberative processes with informally developed public opinions. (p. 298)

Habermas (1992/1998) identifies organisations emerging from both the public and civil sphere as important to public opinion formation, and considers civil society as

composed of those more or less spontaneously emergent associations, organizations, and movements that, attuned to how societal problems resonate in the private life spheres, distill and transmit such reactions in amplified form to the public sphere. The core of civil society comprises a network of associations that institutionalizes problem-solving discourses on questions of general interest inside the framework of organized public spheres. (p. 367)

In this way, Habermas (1992/1998) distinguishes between communicatively generated opinion, and administratively formed opinion, and argues that how

these two processes – the spontaneous forming of opinion in autonomous public spheres and the organized extraction of mass loyalty – interpenetrate, and which overpowers which, are empirical questions. (p. 483)

Habermas (1994, 2008) emphasises the process of systems developing and justifying norms. Habermas (1994, 2008) argues that institutions that develop universal norms and laws need to justify them, and through deliberative democracy, respond to how these norms and laws affect citizens in their application. For the context of educational assessment, the composition and formation of formal forums such as curriculum and assessment panels, and committees in central offices, is therefore an area of continued empirical interest. Particularly how the development of these forums is supported through legislative and bureaucratic process.

Justification and application

The reciprocity inherent in discourses of justification and application is a central aspect of deliberative democracy with analogues in educational assessment. Rehg (2011) describes the discourse of justification as relating to those that focus on justifying a general norm for all those affected, and the discourse of application involving those immediately involved in some situation where the norm seeks to apply. For Habermas (1994), “justifications become rational motives for changes of attitude” (p. 11). These rational motives become evident in discourses of application, which provide the forum for reciprocity between system and lifeworld through deliberative democracy.

The motivation to adopt norms is considered by Habermas (1994) not simply a matter of moral theory, or about the quality of deliberations and discourse in making judgements about norms. Instead, motivation to adopt norms is based on the affective psychological development of individuals that is contingent on socialisation and education. Habermas (1983/1996, p. 207) describes universal morality as dependent on a form of life that it meets halfway through a congruence with the socialised citizen. Where the process of developing socially cohesive societies involves reciprocity between the establishment of justifiable universal norms, and the socialised student motivated to adopt those norms through education. Educational assessment is implicated in these processes, as it communicates societal norms, and provides the basis for discourses of application in the processes of teaching and education that motivates students to adopt them.

There is considerable overlap between notions of deliberative democracy and processes of educational assessment. Deliberative democracy is consistent with Mislav, Steinberg, and Almond (2003, p. 4) who describe educational assessment as communicating values, standards and expectations. Both the linguistic and numerical components of educational assessment are important for setting expectations. The linguistic content describes norms around knowledge and skills valued in society, as well as describing norms on how knowledge and skills are expressed and talked about in society. The numerical component of educational assessment then communicates achievement levels and expectations around levels of competence (Cizek, 2012b). These norms, in terms of linguistic content and achievement levels, are then communicated as expectations through intended interpretations consistent with Moss (2016) and accompanied by justifications consistent

with Habermas (1994). This is to suggest that the deliberative processes associated with educational assessment have a role in establishing norms and to garner a commitment to adhere to them.

The process of application of universal norms in local contexts is also consistent with processes of educational assessment through actual interpretation as described by Moss (2016). Cronin (1994) describes discourses of justification as needing to

be supplemented by discourses of application sensitive to relevant, though unforeseeable, features of situations of action; and moral principles are dependent for translation into action on complementary sources of motivation rooted in structures of identity that are the result of socialization into appropriate forms of social life. (p. xxvii)

Habermas (1994) argues that “the *application* of norms calls for argumentative clarification in its own right” (p. 13). These processes relate to giving feedback to students based on educational assessment (Hattie & Timperley, 2007; Lipnevich et al., 2016). In this way, Habermas’ (1994) processes of justification and application have strong analogues in educational assessment and the process of teaching in education.

Education and law

Habermas (1992/1998) addresses moral norms and laws when developing the concept of deliberative democracy. In this conceptualisation, Habermas (1992/1998, pp. 196, 360) regularly describes education as a background activity that only emerges into the foreground and into law in cases of conflict. However, Habermas (1992/1998) also describes the law as integrating the function of education into broader society, in a similar way to money.

The relationship between the law and educational assessment continues to grow in knowledge-based economies (Drucker, 1968/1992). In the case of Victoria, Australia, the *Education and Training Reform Act 2006* integrates the function of education with other spheres of society. There is also the *Australian Qualification Framework* (AQF) (AQFC, 2013), which comprehensively integrates qualifications ranging from secondary school certificates to doctoral degrees. The framework is “legislated within Australian jurisdictions” (p. 9), and thereby provides an Australian example of how education is

becoming increasingly integrated at the system level and foregrounded through law. The AQF can be characterised as a form of system integration that responds to increasing demands for social integration in schools which in turn respond to increasing social differentiation generated by technological and social progress.

Habermas (1992/1998) develops deliberative democracy and related concepts such as justification and application (Habermas, 1994) to address the legitimacy of laws and norms. These concepts can be extended to educational assessment by linking Kane's (2006) approach to validation to Habermas' (1983/1996) discourse ethics. It is through deliberative democracy that educational assessment can be legitimised in a similar way to how laws and norms are legitimated to meet the needs of society and its citizens.

5.7. Rejoinder to postmodernism

Habermas (1981, p. 13), in an article titled *Modernity versus Postmodernity*, differentiated himself from those he considers as *Young Conservatives*. Habermas (1981) characterises these conservatives as claiming

as their own the revelations of a decentered subjectivity, emancipated from the imperatives of work and usefulness, and with this experience they step outside the modern world. (p. 13)

Habermas (1997) makes similar claims in a later article titled *Modernity: An Unfinished Project*. Both articles are in response to authors such as Derrida (1966/2007), Foucault (1975/1991), and Lyotard (1979/1984) who, as described in Chapter 1, are associated with postmodern critique (Eagleton, 1983/2008; Harcourt, 2007; Leitch, 1992). Where Habermas (1981/1985, 1983/1996) conceptualises the decentered subject through intersubjectivity and discourse ethics, Habermas (1981, 1997) characterises postmodernism as claiming the decentered subject through the subject itself.

As described in Chapter 1, the concept of performativity provides one example where Ball (2003), in the postmodernist tradition, claims the concept for himself to answer his own rhetorical question on "What do I mean by performativity?" (p. 216). In contrast, this study, pragmatically engages with Austin (1962/1975) and Lyotard (1979/1984) as developers of the concepts in an attempt to enhance conceptual clarity. These different attitudes exemplify the different approaches to the decentered subject. Pragmatism

considers the decentred subject is intersubjectively and dialectically established, and for postmodernism, the decentred subject is subjectively established. Postmodernism's focus on the subject raises questions over how it might inform educational assessment design and use, where educational assessment is axiomatically an intersubjective activity.

The universal

This study does not seek to recontest the case between Habermas (1981) and postmodernism, but it does identify justified critiques of modernism from the perspectives of sexuality, gender and race (Butler, 1990/2007; Crenshaw, 1991; Foucault, 1984/1988, 1976/1990, 1984/1990; Irigaray, 1985; Said, 1978/1994; Spivak, 1985/2010; Wittig, 1993). The field of educational assessment is able to address these concerns through Habermas' (1983/1996) principles of universalisation and discourse. It is in entertaining the possibility of developing intersubjectively shared notions of the universal that pragmatism is distinguished from postmodernism.

Habermas (1976/1979, p. 14) works towards intersubjectively developed universal principles, or towards species-wide competence. The validity of these approaches towards universal principles was challenged in a seminal lecture delivered by Derrida (1966/2007) titled *Structure, Sign, and Play in the Discourse of the Human Sciences*. This lecture provided a template for subsequent challenges towards universal principles. In that lecture, Derrida (1966/2007) raises the idea of a rupture caused because the "structurality of structure had to begin to be thought" (p. 249). Here, Derrida (1966/2007) is responding to efforts in previous decades to describe universal structures of language and meaning-making (e.g. Barthes, 1957/1993; Carnap, 1962; Chomsky, 1957/2015; Piaget, 1968/1973, 1947/2001; Toulmin, 1958/2003). Derrida (1966/2007) questioned these meaning-making structures on the basis that structural attempts are founded on the paradox of being "within the structure and outside it" (p. 248). Derrida's (1966/2007) challenge is associated with a rising critique of European ethnocentrism which contributed to a bifurcation between postmodernism and Habermas' (1981, 1997) continued pursuit of modernism (Derrida & Caputo, 1997; Rorty, 1995; Spivak, 1985/2010).

The universal in education

Education is generally considered a universal activity which is reflected, for example, in the *Republic* by Plato (1963) which identifies a responsibility for the state towards children. Kant (1803/2009) describes the unavoidable universal nature of education by observing that the “human being is the only creature that must be educated” (p. 253). In the contemporary context the universal nature of education is recognised by all OECD countries through mandated compulsory years of schooling (OECD, 2016a, p. 475). This is to suggest that the claim that education is a universal concern is largely unproblematic.

To enter into a conversation about education, and then deny its universality, can reasonably be considered as a performative contradiction (Habermas, 1983/1996; Watt, 1975). Consistent with the argument that education is a universal species-wide concern, Habermas (1983/1996) argues for universal consensus. In contrast, Lyotard (1979/1984), argues that consensus is a suspect value, and in favour of smaller narratives with circumscribed consensus horizons, as

any consensus on the rules defining a game and the "moves" playable within it *must* be local, in other words, agreed on by its present players and subject to eventual cancellation. The orientation then favors a multiplicity of finite meta-arguments, by which I mean argumentation that concerns metaprescriptives and is limited in space and time. (p. 66)

Lyotard’s (1979/1984) argument for consensus with limited horizons in space and time is a performative contradiction in education when it is considered a universal activity, and particularly in educational assessment. Educational assessment in both design and use requires consensus positions and background understandings for meaning to remain stable across time and space. If the presupposition of stable consensus is not met, then the field of educational assessment becomes something else. First, large-scale educational assessment design and implementation becomes impossible as design requires sustained meaning-making over domain, technical, and operational matters (Adams & Wu, 2002). Further, interpretations of educational assessment become meaningless, as there can be no concept of stabilised meaning between intended interpretation and actual interpretation (Moss, 2016). Lyotard’s (1979/1984) postmodern solution of paralogy is therefore a performative contradiction for the field of educational assessment.

To enter a conversation about education without a strong sense of the universal is likely to be a conversation about something else. For example, Plato (1993) identifies the sophist as a “hired hunter of rich young men...a wholesaler of learning about the soul...a seller of his own learning” (p. 18) (Schiappa, 2013). Plato’s (1993) distinction between education and sophism has analogies with Habermas’ (1981/1985) distinction between communicative action and strategic action, a distinction further explored in Chapter 6 in the context of markets in educational assessment. When education is considered a universal activity, then questions over how intersubjectivity is addressed come to the fore.

Intersubjectivity and equality

This study addresses intersubjectivity through the universal principles of discourse ethics elaborated by Habermas (1983/1996). In doing so, Habermas (1983/1996) works in the modernist or enlightenment tradition that includes the categorical imperative of Kant (1785/2002) as well as the work of Hegel (1807/1977) and Marx (1848/2000) (Berenson, 1982; Bowie, 2003; de Berg & Large, 2012). Hegel (1807/1977, pp. 111-119) explores intersubjectivity through a narrative of “Lordship and Bondage”, which is used to illustrate that self-consciousness exists only to be recognised by another. A concern for intersubjectivity is continued by Marx (1848/2000) in the *Communist Manifesto*, which somewhat draws on Hegel (1807/1977) to develop the notion of class struggle between the bourgeois and proletarians. Marx (1848/2000) goes on to elaborate the struggle between “Freeman and slave, patrician and plebeian, lord and serf, guild-master and journeyman – in a word, oppressor and oppressed, ... in constant opposition to one another” (p. 246). Habermas (1976/1979, p. 9) continues in this tradition to argue that understanding is based on an intersubjective relationship with others. This suggests that intersubjectivity is well addressed in the modernist tradition.

This study identifies compelling arguments from the feminist tradition (Meehan, 1995) that the tradition of Hegel (1807/1977) and Marx (1848/2000) has difficulty accommodating gender and race. Irigaray (1985), in *This Sex Which is Not One*, argues that women have traditionally been primarily regarded for their exchange value: “Mother, virgin, prostitute: these are the social roles imposed on women” (p. 186). Irigaray (1985) is in effect arguing that women having difficulty entering Hegel’s (1807/1977) master-

slave narrative as an “other”. Wittig (1993, p. 107) makes a similar point in *One Is Not Born a Woman*, arguing that

Marxism has denied the members of oppressed classes the attribute of being a subject. In doing this, Marxism, because of the ideological and political power this “revolutionary science” immediately exercised upon the workers’ movement and all other political groups, has prevented all categories of oppressed peoples from constituting themselves historically as subjects... (p. 107)

Braaten (1995) makes a similar argument, that Marx’s (1848/2000) focus on the means of production has an “intrinsic focus on male workers” (p. 140). Butler (1988) expands on these arguments to argue that “gender might be understood as constituted and, hence, capable of being constituted differently” (p. 520). The arguments of Butler (1988), Irigaray (1985), Wittig (1993), and Braaten (1995), provide compelling reasons for caution towards modernist approaches to the universal.

Ingram (2010, p. 229) considers that feminists have good reason to question Habermas’ approach. Habermas (2000) recognises the political struggle for recognising the gender-specific achievements and interests of women as legitimate. He further sees the potential of this struggle to change “the relationship between the sexes along with the collective identity of women, thereby directly affecting men's understanding of themselves as well” (p. 211). Nevertheless, even in this framing, the onus is with women to initiate the struggle.

It is for marginalised identities that the self-legitimizing narratives, or the *petit recit* [little narrative], argued for by Lyotard (1979/1984, pp. 18-23) opens up possibilities for discourse. That is, when marginalised discourses are denied voice in mainstream discourse, the self-legitimizing discourses proposed by Lyotard (1979/1984) become a viable and somewhat inevitable option. That is, discourses addressing feminism (Fraser, 1990, 2013; Irigaray, 1985, 1995; Wittig, 1993), postcolonialism (Gandhi, 1998; Said, 1978/1994; Spivak, 1985/2010), and sexuality (Foucault, 1984/1988, 1976/1990, 1984/1990) need to self-legitimize when excluded from mainstream discourse. Further, when marginalised identities are excluded from discourse related to educational assessment, then validity and legitimacy of any universal claims arising from educational

assessment become compromised. It is for marginalised identities that Habermas' (1983/1996) approach to the universal is particularly important for maintaining the legitimacy of educational assessment.

Poststructuralism in educational assessment design and use

This study does not seek to argue the broader debate on poststructuralist issues but recognises the need to address the valid claims of feminism (Fraser, 1990, 2013; Irigaray, 1985, 1995; Wittig, 1993) and postcolonialism (Gandhi, 1998; Said, 1978/1994; Spivak, 1985/2010). This study considers Habermas' (1983/1996) universal principles as able to address these concerns. The tension between postmodernism and Habermas' (1983/1996) universal tradition leads to a certain symmetry in performative contradictions. The postmodern approach enters into a performative contradiction through its focus on the subject and on little narratives and paralogy (Lyotard, 1979/1984). In contrast, Habermas' (1983/1996) pragmatism enters a performative contradiction when its claim to universality involves excluding sections of society from its discourse. For the field of educational assessment, these positions frame the broader tension around marginalised identities.

Issues related to marginalised identities and inclusion is not new to the field of educational assessment. Hambleton and Pitoniak (2006, p. 451), for example, argue that panellists for standard setting programs should be representative of stakeholders and constituencies. Schmeiser and Welch (2006, pp. 324-325) similarly argue that item writers and test developers be representative of the target cohort in terms of geography, race, and gender. The notion of inclusion is therefore not foreign to educational assessment. Nevertheless, the poststructuralist critiques of Wittig (1993) and Irigaray (1985, 1995) alert to the possibilities that pragmatism has the potential to render certain identities and perspectives silent if it fails to attend to matters of inclusion.

Closing remarks

This chapter has thematised the work of Habermas (1976/1979, 1981/1985, 1992, 1981/1992, 1983/1996, 1992/1998) for the field of educational assessment. It has described how Habermas (1976/1979) reconstructed historical materialism reframes dialectic between science and ethics. From this reconstruction emerge the theories of legitimation crises (Habermas, 1975/2005), communicative action (Habermas,

1981/1985), system and lifeworld, colonisation (Habermas, 1981/1992), discourse ethics (Habermas, 1983/1996), deliberative democracy, and discourses of justification and application (Habermas, 1994, 1992/1998, 2003).

This chapter has characterised schools as institutions that socially integrate students as citizens into society. Social integration responds to social differentiation which arises in advanced economies through division of labour and the emergence of social identities. Social integration is supported by system integration that involves the generation of autonomous system generated symbolic media such as educational assessment.

Three forms of symbolic media were identified. First is ordinary and everyday language, which is used for general communication including communicative action. Second is generalised communication, which condenses meaning and its use requires validity claims to be addressed whenever it is used. Third is steering media, which Habermas (1981/1992) restricts to money and administrative power and which not just condenses ordinary language but replaces it so that it can be used empirically.

The role of schools in social integration implicates them in motivation crises that emerge from the broader lifeworld of students. Schools are also implicated in legitimation crises that emerge from the relationship that schools have with other systems of society. Legitimation crises emerge when the autonomous sphere of system integration shear off from activities in the lifeworld of students.

This chapter also described communicative action as addressing the objective, social and subjective worlds simultaneously. The theory of communicative action draws on the approaches to argument developed by Toulmin (1958/2003), an approach that Habermas (1983/1996) considers useful but not sufficiently normatively anchored. Habermas (1983/1996) addresses this limitation through a principle of universalisation (U), and a principle of discourse (D). A similar argument can be applied to educational assessment so that validation (Kane, 2006) and design (Mislevy, Steinberg, & Almond, 2003) is similarly anchored in universal principles.

Habermas (1983/1996), following Toulmin (1950/1970, 1958/2003), argues that the structure of scientific arguments and ethical arguments are analogous; scientific arguments draw on perceptions, and ethical arguments draw on feelings. The validity of

both scientific arguments and ethical arguments are established through substantive arguments as described by Toulmin (1958/2003). This study continues to explore how science and ethics are addressed in educational assessment validity, with a particular emphasis in Chapter 7.

This chapter closed with a rejoinder to postmodernism and questions its capacity to inform the design and use of educational assessment. However, in doing so, a potential fallibility in the pragmatic approach is exposed when educational assessment design and use claims to have universal validity while it excludes sections of society from its discourse.

Chapter 6. Educational assessment for economies and politics

This chapter addresses the three purposes of education and educational assessment elaborated in Chapter 4 and how these purposes relate to other autonomous spheres of society. These relationships are developed through theoretical conceptualisations developed by Parsons and Smelser (1956/2005) as well as Habermas (1981/1992). The first section describes education and educational assessment as relating to the economy and polity through several exchanges, mainly: wages for labour; demand in exchange for goods and services; and loyalty in exchange for valid norms and expectations.

The second section makes the case that the exchanges between the household and the economy, and hence the school and economy, are dynamic due to progress narratives in the economy and society more broadly. Parsons and Smelser (1956/2005) identify an integrative system to coordinate these dynamics across systems in society. This chapter elaborates two distinct attitudes towards this integrative system, one in the pragmatic tradition that considers it a function of government, and the second in the neoliberal tradition that considers it a function of markets.

The third and final section of this chapter shows how markets can provide system coordination as advocated by the neoliberal tradition and in a way that relieves the burden on pragmatic reasoning. However, markets are also shown to affect the process of social integration from a universal orientation towards a balkanised one. Three consequences of societal coordination through market-based educational assessment are explored to illustrate how neoliberal approaches to educational assessment can affect society and the lives of citizens.

6.1. The six markets of the AGIL scheme

This section conceptualises the relationship between education and broader society through a schema developed by Parsons and Smelser (1956/2005). Parsons and Smelser (1956/2005) conceptualise education within the *latent pattern maintenance* system of society that is concerned with cultural reproduction. Habermas (1992/1998) conceptualises education in a similar way, where systems

like religion, education, and the family become associated with general reproductive functions of the lifeworld (that is, with cultural reproduction, social integration, or socialization). (p. 360)

Parsons and Smelser (1956/2005) consider the latent pattern maintenance system as having reciprocating relationships with other systems of society. Habermas (1981/1992) largely rejects this approach, and instead conceptualises these relationships through a system-lifeworld model as described in Chapter 5. However, the functional relationships across systems in society are somewhat lost in Habermas' (1981/1992) framing, making it necessary to revisit the structural approach of Parsons and Smelser (1956/2005) for this study. Parsons and Smelser (1956/2005) provide a more useful framework for exploring the interchanges between the different systems of society, and for exploring how educational assessment relates to these interchanges.

The AGIL scheme

Parsons and Smelser (1956/2005) conceptualise a functionalist-structuralist model of society through four systems. This model is based on a biological metaphor that considers systems as successfully performing four functions in a coordinated fashion for survival. These four functional imperatives are: adapting to change, setting goals, integrating change across the system, and maintaining patterns (Heiskala, 2007, pp. 247-248). Parsons and Smelser (1956/2005) respectively describe these functional imperatives through the systems of the economy (A – adaptive), polity (G – goal orientation), integrative system (I), and pattern maintenance (L – latent pattern maintenance). This scheme is referred to here as the AGIL scheme (Chernilo, 2002; Heiskala, 2007; Holton & Turner, 1986). These four functional imperatives led Parsons and Smelser (1956/2005) to identify six boundary interchanges, or six markets, which are explored here in the context of educational assessment. A representation of the AGIL scheme with its six markets is illustrated in Figure 8.

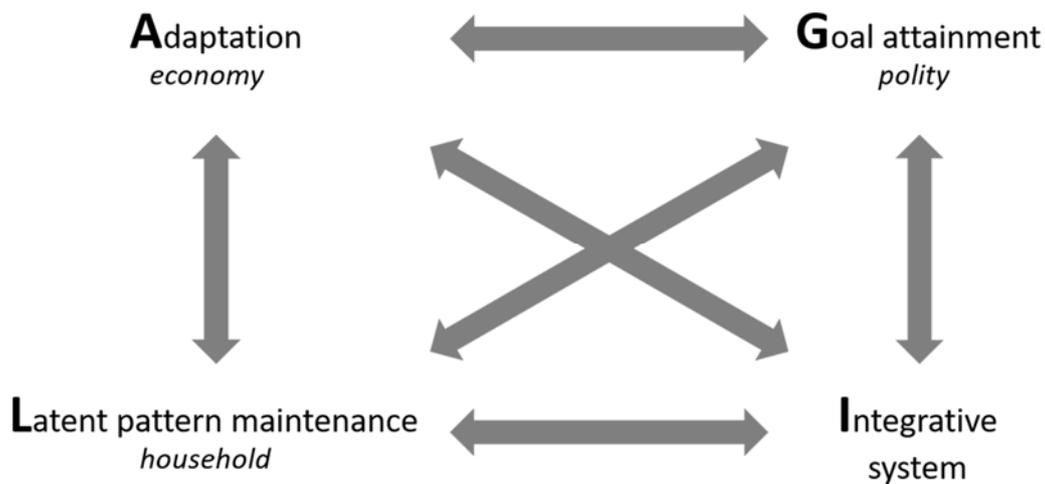


Figure 8 – The AGIL system and six markets
Parsons and Smelser (1956/2005)

The exchanges conceptualised by Parsons and Smelser (1956/2005, p. 67) contrast to the system-lifeworld framing described in Chapter 5. Habermas (1981/1992) describes his reframing as one where

Parsons held that all systems of action constitute environments for one another, develop their own media, and regulate intersystemic interchange via these media; by contrast, our two-level concept of society requires that we distinguish between the perspectives of system and lifeworld. (p. 319)

This study pragmatically adopts the approaches of both Habermas (1981/1992) and Parsons (1963a) finding that each provides a useful framing for educational assessment. Following Habermas (1981/1992), educational assessment as a medium of influence can be normatively anchored in the lifeworld. This study considers the lifeworld somewhat analogous to the latent pattern maintenance system in the framing of Parsons and Smelser (1956/2005). However, this chapter follows Parsons (1963a) to provide a better framing for the increased demand for institutionally anchored media in knowledge-based economies (G. S. Becker, 1964/1993; Drucker, 1968/1992; Hanushek, 2016; Hanushek & Ettema, 2017; Levy & Murnane, 2004; Newton, 2007). This study conceptualises educational assessment as a media of influence across autonomous systems of society with a focus on the economy and polity.

The framings of both Habermas (1992/1998) and Parsons and Smelser (1956/2005) do not foreground education which results in both framings not differentiating the school from the family. This is an issue for this study, and this section foregrounds education, and the role of educational assessment for the respective models. While Parsons and Smelser (1956/2005) argue that all systems of interaction constitute environments for one another, and that each develop their own media, they nevertheless do not address educational assessment as media. Furthermore, as previously described, Habermas (1981/1992) argues against the overgeneralisation of media, to argue that

the private and public spheres are communicatively structured spheres of action, which are not held together by systemic means—that is, not by steering media... (p. 319)

But this study observes that educational assessment can be institutionally anchored to provide means of coordinating systems such as school funding decisions (Gonski, 2011; Shine, 2016). Further, educational assessment mediates between the private sphere of the household and the public sphere of the economy and polity through teachers providing student reports to families (Kohl, Lengua, & McMahon, 2000). This study will therefore proceed on this basis, and contrary to Habermas (1981/1992), in conceptualising educational assessment as media for systemic and intersystemic communication as well as for communication within the lifeworld.

While Parsons (1963b) conceptualises a range of symbolic media for the AGIL scheme – including money, power, influence and commitment – this study will not address these media. Instead, this study focuses solely on conceptualising educational assessment as symbolic media in the AGIL scheme. Of particular interest is that Parsons and Smelser (1956/2005) consider the latent pattern maintenance system of AGIL scheme as having a “special boundary” with other systems in society.

The cultural system and latent pattern maintenance

Parsons and Smelser (1956/2005, pp. 68-69) associate the latent pattern maintenance system with aspects of the household and as having cultural boundaries rather than interaction boundaries. Parsons and Smelser (1956/2005, p. 9) conceptualise a performance-sanction paradigm to describe interactions between different systems. A *performance* provides a “contribution” to the functioning of the system, and a *sanction*

describes the effect on the system towards which a performance is directed. This study finds that the performance-sanction paradigm, when considered from the perspective of educational assessment, is analogous to the intended interpretation and actual interpretation model described by Moss (2016). This study uses the intended interpretation and actual interpretation model of Moss (2016) to refer to the use of educational assessment across autonomous systems of the AGIL scheme as well as for communication within the latent pattern maintenance system itself.

Parsons and Smelser (1956/2005, p. 69) describe the latent pattern maintenance system as insulated from the performance-sanction interplay of the larger system. For Parsons and Smelser (1956/2005), the “special” boundary of the latency system is a cultural rather than an interaction boundary. The

latency sub-system ... maintains value patterns. But cultural patterns are not isolated atoms, each institutionalized in connection with its own particular system or sub-system. The cultural value system of a society is more or less integrated. In particular the value patterns applicable to any given sub-system are *differentiated* value *sub-systems* of the general value system of the total society (p. 69)

The latent pattern maintenance system, and its analogue of the lifeworld, is of interest to this study in terms of the relationship between culture and other systems of society through social integration as a function of schools. The function of schools in the AGIL scheme is next elaborated before the integrative system in relation to schooling is explored.

The school

The AGIL scheme of Parsons and Smelser (1956/2005) is highly abstract and does not presuppose any particular organisational form. It does not

correspond directly to concrete organizational units. Its relation to organization varies from society to society and in the same society over time. (p. 57)

For the purposes of the present study, the concrete notion of a school is postulated as the organisational unit for education. Elmore (1996), for example, considers schools as

having classrooms with an *instructional core* of students, teachers and content. The school remains a clear and stable concept for fulfilling the function of education in society. This is despite developments in technology, and periodic attempts to reconceptualise education typified by Illich (1970/1973) in his seminal book *Deschooling Society*. The concept of a school and its location in the AGIL scheme is illustrated in Figure 9.

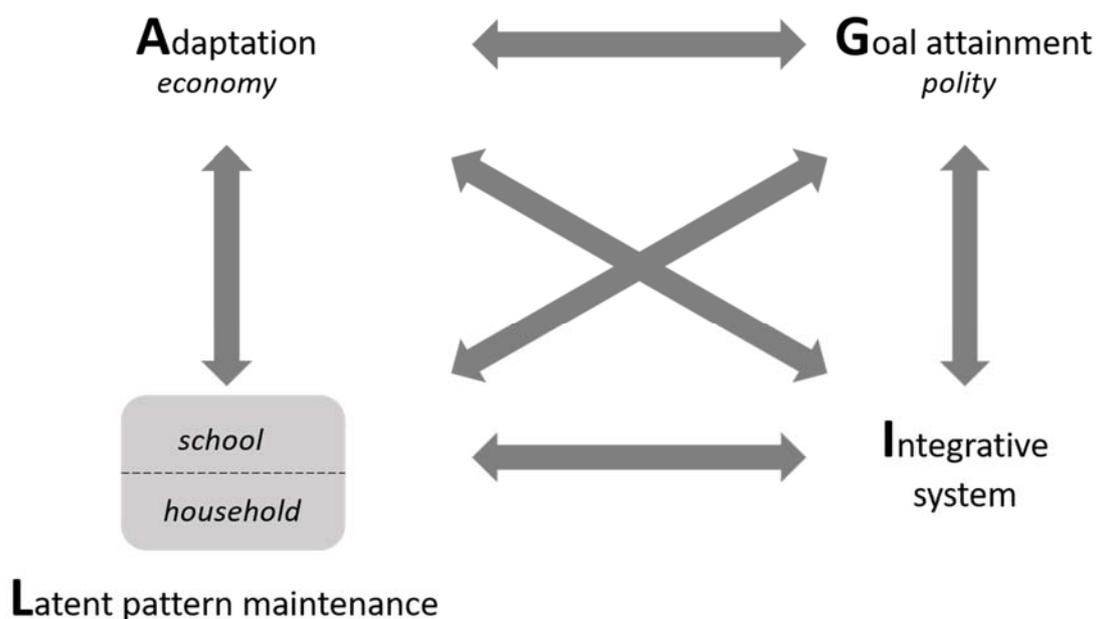


Figure 9 – The AGIL system with school differentiated from household
Parsons and Smelser (1956/2005)

The differentiation between school and household illustrated in Figure 9 results from an increasing functional relationship that education has with both the economy and polity in advanced and knowledge-based economies (Drucker, 1968/1992). Here, the concept of function is used to describe what society asks of education to produce for the economy and polity, and not to describe how a school functions (Durkheim, 1933/2012, p. 49). Education is increasingly conceptualised as being linked to the economy in knowledge-based economies through human capital theory (G. S. Becker, 1964/1993; Drucker, 1968/1992; Hanushek, 2013; Hanushek et al., 2013; Hanushek & Wößmann, 2010; Levy & Murnane, 2004). Further, through increased social differentiation the school is being drawn out of the cultural background to become increasingly differentiated from the home as described in the latent pattern maintenance system of Parsons and Smelser

(1956/2005), and the lifeworld of Habermas (1981/1992). These dynamics expose borderline cases of how schools relate to both the family and society more broadly.

The six boundary exchanges are described by Parsons and Smelser (1956/2005) and will not be explored here in similar detail. Instead, the focus will be on how educational assessment is implicated in exchanges between systems, particularly in exchanges involving the latent pattern maintenance system, the economy and polity. Of further interest is how these exchanges relate to the integrative system to accommodate societal progress along technological and social dimensions.

Educational assessment is implicated in the AGIL scheme in the following ways, for which completeness is not claimed. First, through qualification and certification, educational assessment provides symbolic media between the household and the economy on matters of employment. When taste is considered as malleable as described in Chapter 4, education and therefore educational assessment becomes implicated in taste formation which relates to the nature of demand in the economy. Second, there are two relationships with the polity. The first relates to the legitimacy of institutionally anchored educational assessment governed by the polity, the second relates to the polity garnering loyalty towards societal norms. The latter relationship is explored, for example, by the assessment of civics and citizenship (W. Schulz, Ainley, Fraillon, Losito, & Agrusti, 2016). Of central interest however is the relationship between the latent pattern maintenance system and the integrative system. This study finds that in societies advancing technologically and socially, the content of educational assessment needs to evolve to reflect these developments. How these developments in society are mediated to education and educational assessment is explored next in relation to ideologies.

6.2. The integrative system – pragmatism and neoliberalism

The previous section described how education, and hence educational assessment, can be conceptualised as oriented towards the economy and polity. This section specifically explores the role of the integrative system and how it mediates between the latent pattern maintenance system and other systems of society. In particular, the distinctive attitudes of neoliberalism and pragmatism towards the integrative system are explored. To illustrate the kind of dynamics an integrative system might be required to address in education, three selected tensions are explored.

The three tensions are illustrative, and no claim is made to completeness that these describe the full range of interactions that schools have with broader society. The three selected tensions address how the school is implicated in stresses between: the public and private spheres (Fraser, 2013; Friedman, 1962/2002; Habermas, 1962/1991), the maintenance and fluidity of wants (Friedman, 1962/2008; Peters, 1959/1973), and education for the economy (work and vocation) and education for the polity (citizenship) (Friedman, 1962/2002; Parsons & Smelser, 1956/2005). How these three tensions manifest in schools is largely socioculturally determined and can be conceptualised as being mediated in society by an integrative system.

While the integrative system is explored later from both the perspectives of pragmatism and neoliberalism, the selected tensions themselves do not emerge from this ideological divide. One selected tension emerges from poststructural critique of gender and sexuality. The three tensions are described next before distinctive attitudes towards the integrative system are explored.

The school as public or private sphere

The relationship that schools have with the public and private sphere, or more particularly the relationship that the state's administrative structures have with the private sphere, can vary. This study observes Habermas' tradition and neoliberalism as fundamentally sharing similar and traditional views on this division. The division between the public and the private is instead challenged from the feminist perspective (Fraser, 1990), and from emerging issues involving sexuality (Rodwell, 2016).

Friedman (1962/2002), in a neoliberal tradition, considers parents as best placed to protect children, where the "ultimate operative unit in our society is the family" (p. 33). This view is somewhat shared by Habermas (1962/1991) in his analysis described in *The Structural Transformation of the Public Sphere*, which details the evolution of the demarcation between the public and private spheres. However, both Friedman (1962/2002) and Habermas (1962/1991) recognise a level of ambiguity. For Friedman (1962/2002),

children are at one and the same time consumer goods and potentially responsible members of society. The freedom of individuals to use their economic resources as they want includes the freedom to use them to have

children to buy, as it were, the services of children as a particular form of consumption. But once this choice is exercised, the children have a value in and of themselves and have a freedom of their own that is not simply an extension of the freedom of the parents. (p. 33)

Similarly, Habermas (1981/1992) considers the child as having basic rights which requires the possibility of state intervention, he also observes that

one can understand the policy recommendation to the effect that legislators keep to a minimum the state interventions necessary to protect children's rights. Legislative regulation, therefore, ought not to favor far-reaching judicial intervention ... (p. 370)

Fraser (2013) provides a contrasting feminist perspective that challenges the positions of both Friedman (1962/2002) and Habermas (1981/1992), by arguing that the division between the private and public is

discursively constructed, gender- and power-saturated objects of political struggle; and I link the politicization of needs to feminist struggles over where and how to draw the boundaries between “the political,” “the economic,” and “the domestic.” (p. 7)

Where both Friedman (1962/2002) and Habermas (1981/1992) indicate a preference for a division between private and political life, for Fraser (2013), “the personal is political” (p. 1).

The argument that Fraser (2013) makes in regard to the public and private division, and how this affects women differentially, particularly those Fraser (2013) describes as childrearing, is a clear case. In the state of Victoria, for example, state intervention into the private sphere of the home is authorised through the *Child Wellbeing and Safety Act 2005* which mandates home visits by state-sponsored maternal health nurses. These visits authorise state actors, in the form of maternal health nurses, to check on the private lives of mothers on matters such as smoking, family health, family violence, car safety, and psychological health (DEECD, 2011, p. 51). This is to provide clear case support to the argument by Fraser (2013) that for mothers the personal is political and the personal and the intimate is of interest to the administrative function of the state.

A child's legal relationship to its mother and broader family, and how this relationship comes under the purview of the state from birth, illustrates how schools reach into both the public and private spheres. While a child's individuation from its father can remain in the private sphere in a manner consistent with the position of Friedman (1962/2002) and Habermas (1962/1991), a child's relationship to its mother is of a different character. A child's individuation from its mother involves state actors from birth. It first involves legislated maternal health nurses and then legislated compulsory schooling (DEECD, 2011, p. 51; OECD, 2016a, p. 475). This renders the mother-child relationship inherently political and subject to the public sphere.

Further contestation around the public and private boundary emerges around sexuality. For example, the education system in Victoria has a *Safe Schools* program that addresses bullying and harassment that arise from student sexual identities (Department of Education and Training Victoria, n.d). However, these programs are challenged by parent groups (Rizmal, 2017), with Rodwell (2016) observing moral panic arising from certain quarters of society. Several phenomena are illustrated here. There is social differentiation at the sociocultural level on matters of sexual identity, with the moral panic observed by Rodwell (2016) illustrating a motivation crisis in the sociocultural system. Issues of sexuality illustrates the tensions described by Friedman (1962/2002) where a child is both a parental consumer good and having a value for themselves with freedoms requiring state protection. Contested notions of sexuality provide further support for schools reaching into both the public and private spheres.

This study does not seek to develop a position on how the public and private boundaries should be implemented in schools, but only seeks to show that a tension exists along this boundary. Further, this study observes that this tension implicates an integrative system to manage dynamics around this boundary that might emerge from progress narratives in society.

The school as influencing wants

This study explores Friedman's (1962/2008) assumption that wants are fixed and upon which Friedman's (1953/2008, 1969/2008, 1962/2008) economic models are based. The concept of education as initiation developed by Peters (1959/1973) will be used to

juxtapose Friedman's (1962/2008) assumption to expose a role for the integrative system to manage the fluidity of wants.

Peters (1959/1973) develops a comprehensive and nuanced framework for what education and teaching entails that is further described and critiqued by Cuypers and Martin (2013, 2011). The interpretation provided here will therefore be reduced to focus on how it pertains to the argument. Peters (1959/1973) works in a pragmatic tradition that has a focus on ethics and intersubjectivity. In a similar fashion to Habermas (1976/1979), Peters (1959/1973) considers cognition and skill as important aspects of education, but that these be accompanied by an approach to ethics that places these skills in context and within "a coherent pattern of life" (p. 61) (C. Martin, 2012).

Peters (1959/1973) cautions against teaching approaches that neglect the fluidity of wants, and argues that the

job of the educator is not simply to build on existing wants but to present what is worth wanting in such a way that it creates new wants and stimulates new interests. (p. 105)

Peters (1959/1973) goes on to argue that if teachers do not address the development of wants in students, others such as advertisers and peers will. The positions of Peters (1959/1973) and Friedman (1962/2002) expose a tension that is conceptualised here as being managed by an integrative system.

The school as education or vocational education

The relationship that a school has with the economy and polity presents another point of tension. Friedman (1962/2002) articulates this tension well when he claims that there are

different implications for (1) general education for citizenship, and (2) specialized vocational education. The grounds for governmental intervention are widely different in these two areas and justify very different types of action. (p. 86)

In the context of the AGIL scheme, a school's relationship to the economy through vocational pathways is distinct to a school's relationship to the polity through education for citizenship. The tension between these two purposes affects how educational assessment is considered. Newton (2007), for example, considers secondary school

certification as providing an indication that a student is “at least minimally qualified, if only for an aspect of everyday citizenship” (p. 162). Newton (2007) is here implying multiple orientations for secondary school exit certificates, while suggesting that they are more tightly bound to citizenship than vocational pathways.

The distinction made by Friedman (1962/2002, p. 88) has more significance through his related argument that government is only justified in subsidising education for citizenship and not justified to subsidise vocational education. While Friedman (1962/2002) acknowledges that it is difficult to distinguish between the two forms of schooling, the general distinction has had effect in policy. As described in Chapter 1, the distinction between education for citizenship and education for vocation led to the introduction of a university fee scheme of a style advocated by Friedman (1962/2002) in Australia during the 1980s (Laming, 2012; Marginson, 1993).

Further, while Friedman (1962/2002) could be considered here as describing a philosophical position, the distinction relates directly to Friedman’s (1969/2008) monetarist approach to economics. Friedman (1969/2008, pp. 48-50) postulates four kinds of capital for his economic models: physical productive, human productive, physical consumption and human consumption. The economic models postulate that income received from an individual’s physical and human productive capital can be reinvested to increase human productive capital to further enhance its return. In this economic model, vocational education is considered a property and asset of the citizen where the citizen is responsible for the development of this asset. That is, a student’s economic productive capacity is not a property or responsibility of society in Friedman’s (1969/2008) model, but the responsibility of the citizen. In the context of the four types of capital postulated by Friedman (1969/2008), this claim is an empirical claim as well as a philosophical one (Friedman, 1962/2002).

A tension therefore exists between education for citizenship and education for vocational education. This tension is somewhat exacerbated by legislated compulsory schooling across the OECD (2016a, p. 475) that extends to 18 years of age in some countries. Within this age range, many students start to prepare for, or commence, vocational pathways. In Victoria, for example, education is compulsory up to and including 17 years of age during which vocational pathways are encouraged and pursued (Education and Training Reform

Act 2006). Given that school attendance is mandated by the state through legislation, there is weight to the argument that the state has a responsibility for ensuring that students are offered appropriate vocational pathways.

Given that vocational pathways are pursued during the compulsory years of schooling, there is an onus on the government to provide advice to students as well as an onus to make appropriate training pathways available. This implicates an integrative system to mediate between how work is carried out in the economy and what vocational skills are delivered during the compulsory years of schooling. The imperative for this advice increases for dynamic economies where technological progress might affect the nature of work.

Neoliberalism and pragmatism

How the three identified selected issues manifest is socioculturally and politically determined and is also a function of technological and social progress. This study conceptualises the integrative system of the AGIL scheme as responsible for establishing and maintaining patterns across autonomous systems of society pertaining to these tensions. Neoliberal and pragmatic approaches will now be used to explore different attitudes towards the integrative system.

To better frame the differences between neoliberalism and pragmatism towards the integrative system the term individualism will be used here instead of neoliberalism. Individualism is largely synonymous with neoliberalism and is a term adopted by both Hayek (1944/2007) and Popper (1945/2002). Hayek (1944/2007) frames his argument through individualism to illustrate how neoliberalism places greater value on the rational behaviour of individuals than the coercive deliberations of the collective (Friedman, 1962/2002; Popper, 1945/2002; von Mises, 1944/2007). It is the tension between collective decision-making and individual freedom that is used here to explore different attitudes towards the integrative system.

Neoliberalism takes a strong position on the role of government in education, where a

certain amount of state control in education, for instance, is necessary, if the young are to be protected from a neglect which would make them unable to defend their freedom, and the state should see that all

educational facilities are available to everybody. But too much state control in educational matters is a fatal danger to freedom, since it must lead to indoctrination. (Popper, 1945/2002, p. 106)

A similar argument towards education is advocated by Friedman (1962/2002). However, while arguing in favour of compulsory years of schooling, both Popper (1945/2002) and Friedman (1962/2002) leave under-conceptualised what content should be taught during the compulsory years of schooling. Friedman (1962/2002, p. 89), for example, invokes concepts such as “approved educational services”, “approved institutions” and “approved school”. However, Friedman (1962/2002) does not conceptualise how criteria for approval should be developed and administered within complex plural societies.

The neoliberal attitude towards education described here through Popper (1945/2002) and Friedman (1962/2002) contrasts to pragmatic approaches. Where neoliberalism seems largely unconcerned over what constitutes approved educational content, this question is at the heart of Habermas’ (1981/1985, 1981/1992, 1983/1996) pragmatics that addresses the validity of norms. Plato’s (1963) *Republic* provides a quintessential example of pragmatic deliberations over educational content to also provides an example with which Popper (1945/2002) explicitly raises concerns in the neoliberal tradition.

The contrasting attitudes that neoliberalism and pragmatism bring to education can be considered through Figure 10, which is a reformulation of the AGIL scheme to highlight the role of the integrative system in education. Figure 10 illustrates the relationship that both the economy and polity have with the sociocultural system of the household and school.

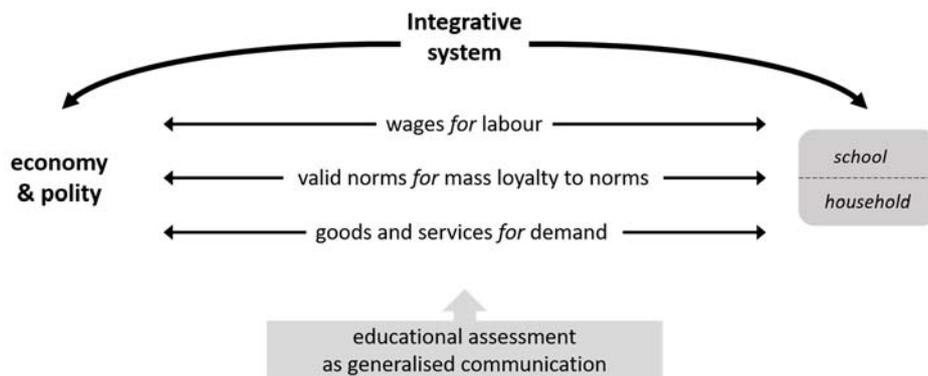


Figure 10 – Educational assessment and the integrative system

The key exchanges that the polity and economy have with schooling draws directly on the framing of Parsons and Smelser (1956/2005) as well as Habermas (1981/1992, p. 320). These relationships also relate to the selected tensions described earlier. This study makes the case that these exchanges are supported through the symbolic media of educational assessment. Education and educational assessment support the wages for labour exchange through qualification, the loyalty to norms in exchange for valid norms through socialisation, and the demand for goods and services in exchange for demand through development of the self. There is therefore direct correspondence between these exchanges and the purposes of education and educational assessment described in Chapter 4.

However, the key argument addressed here relates to how pragmatism and neoliberalism consider the integrative system. A pragmatic approach in the tradition of Habermas (1992/1998) might consider the integrative system a function of deliberative democracy where government has a responsibility to provide guidance to students on skills required for the economy during the compulsory years of schooling. This approach might also postulate a role for government in establishing, communicating and justifying norms that are universally valid and worthy of adherence by students as citizens. Government might further have a role promoting, supporting and protecting appropriate forms of self-expression by students as they emerge as citizens.

As illustrated in Figure 10, how the responsibilities of government might be pragmatically fulfilled depends on broader dynamics including technological and social progress. That is, the kind of labour supplied to the economy is dynamic, the goods and services demanded by citizens from the economy is similarly dynamic, and the expectations that society has of its citizens is dynamic.

Where the pragmatic tradition might seek to integrate progress narratives into educational assessment through government institutions that respond through deliberative democracy, the neoliberal tradition and individualism eschews institutional interference. Instead, neoliberalism favours markets. By advocating for a reduced role for government in education (Popper, 1945/2002), neoliberalism turns away from managing complexity emerging from progress narratives through language. In this way, neoliberalism turns away from communicative reason in the face of increasing complexity and assigns

responsibility for resolving emerging issues to markets. This renders vague the status of educational content relating to how work is performed in the economy and the kind of norms society expects students to adhere to as they emerge as autonomous citizens. By eschewing an active deliberative interest in educational content this content potentially becomes reified through an unarticulated status quo or drift to take a path autonomous of broader societal expectations. By not actively addressing broader dynamics in society neoliberalism has the potential to exacerbate motivational crises around issues such as gender and sexuality (Fraser, 2013; Habermas, 1975/2005; Rodwell, 2016).

6.3. Markets as strategic action

Strategic action was described in Chapter 5 as social action that is goal-oriented (Habermas, 1981/1992). This study finds that the field of marketing, as described through the work of Kotler (1963, 1964, 1965, 1972), is one form of strategic action that bears on educational assessment design and use.

In characterising marketing as a form of social action as described by Habermas (1981/1992), markets can be shown to reduce the requirement for postconventional moral reasoning (Habermas, 1983/1996; Kohlberg, 1971). Markets reduce the burden to reach consensus positions among a plurality of views and values. However, while relieving the burden on pragmatism and language to develop consensus and common understandings, markets also affect processes of social differentiation and social integration.

Marketing as social action

A relationship between strategic action, as described by Habermas (1981/1985), and the field of marketing becomes formally evident through axioms of marketing developed by Kotler (1972). Kotler (1972) identifies four axioms as follows

Axiom 1. Marketing involves two or more social units, each consisting of one or more human actors. ...

Axiom 2. At least one of the social units is seeking a specific response from one or more other units concerning some social object. ...

Axiom 3. The market's response probability is not fixed. ...

Axiom 4. Marketing is the attempt to produce the desired response by creating and offering values to the market. (pp. 49-50)

Kotler's (1972) axioms mesh with the concept of strategic action developed by Habermas (1981/1985). The first, second and fourth axioms declare marketing as social action that is goal-oriented consistent with Habermas' (1981/1985) definition of strategic action. The third axiom is not hostile to Habermas' (1981/1985) definition and provides an additional characteristic that defines marketing as a specific form of strategic action further explored here.

Kotler (1972) expands the third axiom with an important corollary that "if the target social unit cannot respond to the social object, as in the case of no interest or no resources, it is not a market" (p. 50). This corollary provides a way of distinguishing an actor as citizen and an actor as being in market. That is, if a citizen has no interest or resources to respond to an offered product they are not in the market for that product.

A further corollary made by Kotler (1972) is that a "market's response probability can be altered by marketer actions" (p. 50). That is, a person marketing a product can tailor a product definition to enhance the probability of a desired response. The tailoring of products and messages leads to marketing concepts such as target markets, market segmentation, product definitions, and differentiated marketing strategies (Kotler & Levy, 1969).

The marketing axioms of Kotler (1972) illustrate that a market is not a society of citizens, and instead a subset of the citizenry, or subset of society, that shares an interest with a marketer and with the resources to pursue those interests. By providing a mechanism through which like-minded social actors can exchange goods and services, markets lower ethical demand by relieving the requirement to address complexity arising from cultural plurality. In this way, markets affect processes of social differentiation and social integration.

In providing actors with an opportunity to engage with like-minded actors, marketing relieves ethical demand. Marketing allows actors to avoid the kind of moral development required to engage with principles that extend beyond their own group and interests. In terms of Kohlberg (1971, pp. 86-87), marketing allows social actors to engage at

conventional levels of moral development that simply requires conformity and loyalty. Targeting like-minded social actors avoids the requirement for postconventional reasoning with those who might have different values and attitudes.

Marketing affects social differentiation by allowing social actors to target other actors with similar interests and the resources to engage with those interests. Markets allow citizens to create lifeworlds with those who have similar interests and resources in developing life-projects. There is no inherent need for a social actor to socially integrate with social actors with differing cultural values and views when engaged in markets. In this way, marketing affects social differentiation by facilitating and amplifying differentiation, and relieving the need for social integration.

Consistent with how Habermas (2001) describes postmodernism and neoliberalism, this study also finds that both lead to similar consequences. Where postmodernism has the little narrative and localised consensus limited in time and space (Lyotard, 1979/1984), marketing allows like-minded social actors to negotiate life-projects without the burden of developing attitudes that address a plurality of views or universal views. As Habermas (2001, p. 88) suggests, both positions lead to a vision of a lifeworld scattered into small groups and discrete monads. In this way, educational assessment can affect the life of citizens through how it responds to markets.

Marketing in education

The axioms of Kotler (1972) describe marketing as a process that scatters and balkanises citizens into probabilistic groups that share interests and have the resources to pursue those interests. Marketing also describes the targeting of products to target groups. From the perspective of this study, the variables of class, sexuality, gender and race present themselves as variables around which target markets can be identified. The phenomenon of commodity feminism, as described by Goldman, Heath, and Smith (1991), provides one example of how these variables can be used to conceptualise a market. These possibilities are offered here because the variables of class, sexuality, gender and race are of interest to this study. However, this study does not pursue this line of argument, and instead presents tensions that markets bear on educational assessment. Three tensions are presented.

Markets in educational assessment

Chapter 4 identified pressures on Australia ATAR illustrated by the comments of Australia's Chief Scientist (Finkel, 2018). Finkel (2018) questions the capacity of the ATAR to effectively coordinate a student's transition from secondary to tertiary education, and observes students and universities as increasingly taking a goal-oriented and instrumental attitude towards the ATAR.

The ATAR, as a symbolic media, is best described in terms of generalised communication. The ATAR is a numerical summary of educational activity and academic achievement that reflects activity across a wide variety of subject areas that are delivered across various jurisdictions. A student's ATAR is a product of a student's subject choice in secondary school, the effort expended in pursuing those subjects, and a student's innate capability. The ATAR is an instance of generalised communication that pertains to objective world of qualification, the social world of fair access to tertiary education based on merit, and to subjective world of student's personal aspirations and capabilities. These characteristics are all consistent with the ATAR as a form of generalised communication addressing the pattern of three worlds (Habermas, 1981/1985, 1981/1992; VTAC, 2016).

The purposive and empirically motivated use of the ATAR by both universities and students identified by Finkel (2018) is an indication of the ATAR losing its communicative capacity. In terms of Habermas (1981/1992, 1992/1998), the purposive use of the ATAR is consistent with the ATAR becoming delinguistified in a way that it is no longer able to communicate to universities the subjective career aspirations of the student along with their capacity to pursue those aspirations. This illustrates both a motivation crisis in terms student subject choice, and a legitimation crisis in terms of the ATAR being able meaningfully communicate a student's career disposition across a wider system.

The literature review in Chapter 2 identified broader concerns towards the ATAR, with Blyth (2014) predicting that the ATAR will survive long-term. In contrast, Matters (2015, p. 72), from ACER, predicts that the ATAR will break down over the medium-term to long-term. The position of Matters (2015) is particularly poignant as she also provides advice to Australian governments on secondary school exit credentials (Matters & Masters, 2014). ACER also develops and administers proprietary university admission

tests for medical and law courses (S. L. Elliott & Epstein, 2005; Poole, Shulruf, Rudland, & Wilkinson, 2012). In this sense, ACER is both providing advice to government, while at the same time having a clear interest in the debate as a provider of proprietary media competing with the ATAR. The position of ACER towards the ATAR is therefore problematic given that ACER has a direct interest in the secondary to tertiary education transition.

Tensions around the ATAR are also fuelled by a desire to increase the validity horizons of credentials. As observed by Habermas (2001, p. 83) in *The Postnational Constellation*, modernisation opens up new intersubjectively shared lifeworlds, where the impulse to broaden lifeworlds is generated by new markets, new means of communication, as well as new modes of commerce and cultural networks. This desire is illustrated by the growth in the International Baccalaureate qualification as a competing qualification to, for example, the Victorian Certificate of Education (VCE) in Victoria (Bryant & Walker, 2016). Rahimi, Halse, and Blackmore (2017), in an analysis of the Victorian school sector, describe processes of internationalisation and commercialisation as changing the landscape for Victorian schools. In this sense, internationalisation and commercialisation, and their concomitant impulse of citizens to broaden lifeworld horizons, is placing increasing pressure on the secondary to tertiary education transition.

This study identifies two broad models for educational assessment to steer student transition from secondary education to tertiary education, with each broadly aligning with pragmatic and neoliberal traditions. The first model, in the pragmatic tradition, is exemplified by the ATAR. The ATAR provides motivational support for social integration in that it allows students to pursue different life-projects while maintaining enfranchisement to a universal process of transition from secondary to tertiary education. In the pragmatic tradition, the issues identified by Finkel (2018) with the ATAR might be addressed by improving the ATAR in cooperation with the various jurisdictions, secondary school sectors, and the university sector. This form of problem solving requires communicative action and postconventional reasoning (Habermas, 1981/1985; Kohlberg, 1971).

A second model, more closely aligned to the neoliberal tradition, is to have the transition from secondary to tertiary education mediated by private and commercially generated

symbolic media. This second model leads to markets in educational assessment as symbolic media through proprietary tertiary selection tests (B. Griffin, Bayl-Smith, & Hu, 2018; Mercer, Hay, Hodgson, Canny, & Puddey, 2018; Poole et al., 2012). However, such an approach has the potential to adversely affect student individuation. Through the ATAR, students are encouraged to pursue school subjects that align with personal aspirations. However, when symbolic media is no longer able to incorporate the personal aspirations of students and their subject choice the symbolic media loses the capacity to motivate student individuation.

Given that markets are defined as those with an interest in a product and the resources to engage, markets have a problematic relationship to universal enfranchisement to affect social integration. That is, enfranchisement to market-based symbolic media is affected by a student's financial capacity to participate. There is a significant cost involved in sitting the Undergraduate Medical Admissions Test (UMAT), for example. This cost increases with re-sitting of the test, with a study by Puddey, Mercer, Andrich, and Styles (2014) concluding that "Re-sitting the UMAT augments performance in each of its components together with the total UMAT percentile score" (p. 1). There is also a market for coaching in the UMAT, however a study by B. Griffin, Harding, Wilson, and Yeomans (2008) concluded that coaching does not assist in performance. Nevertheless, in terms of Kotler's (1972) marketing axioms, the financial cost of sitting, and the additional cost of re-sitting and coaching to enhance success, has the potential to exclude some, and advantage others through available resources.

Proprietary tertiary selection tests that compete with, or supplement, established ATAR processes continue to emerge in Australia. Edwards, Coates, and Friedman (2012), in an international survey on behalf of the OECD, identify a range of targeted new university admission tests, including the: UMAT for medicine, Australian Law Schools Entrance Test (ALSET) for law, and Aptitude for Engineering Assessment (AEA) for engineering.

The influence of commercial interests in education has been of continuing concern. Lyotard (1979/1984) raises this concern through the concept of performativity. Lyotard (1979/1984, p. 51), considers that the mercantilization of knowledge shifts the criteria for what becomes educationally important, from "Is it true?" to "What use is it?", "Is it saleable?" and "Is it efficient?". The concerns over the effect of commercial interest in

education and educational assessment are ongoing (Burch, 2009; Burch & Good, 2014; Hogan, 2012, 2016a; Hogan et al., 2015; Marginson, 1997b). These concerns on how public and private values can affect educational assessment as symbolic media are illustrated in Figure 11.

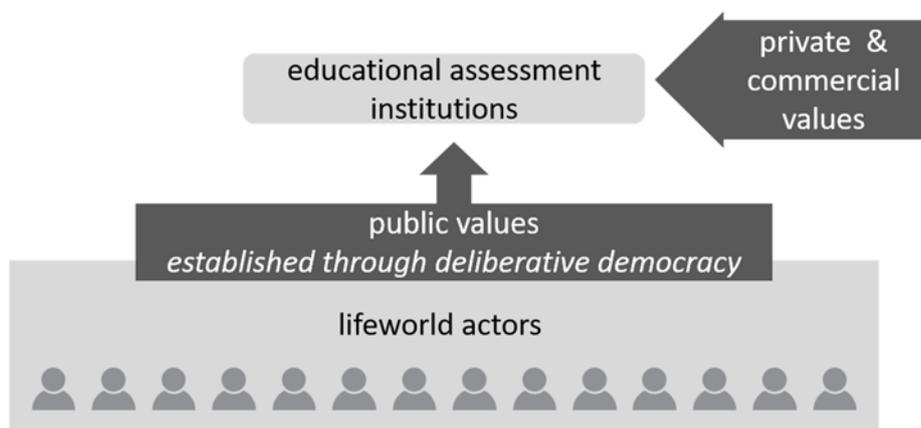


Figure 11 – The influence of public values, and private and commercial values

This study does not seek to identify how tensions around educational assessment that result from increasing social differentiation and internationalisation should be addressed. Instead, this study suggests that an increased focus on market-based and commercially generated educational assessment can work against processes of social integration. On the one hand, market-based educational assessment media can relieve the burden on linguistic communication and postconventional reasoning. On the other hand, in relieving this burden, market-based educational assessment media can work against processes of social integration and motivation to have a balkanising effect on society. These issues pertain to matters of legitimacy, and therefore might be considered in processes of educational assessment validation (Kane, 2006) as part of the integrative system of society (Parsons & Smelser, 1956/2005).

Markets in educational assessment design – the Ultranet

Markets and strategic action can be antagonistic towards communicative action in educational assessment design and use. This is made evident in commentary subsequent to the failure of the Ultranet project in Victoria, a project that “was expected to provide an online teaching, learning and assessment system for all government schools” (Pearson, 2012, p. 20). The transcript of evidence from the resulting corruption inquiry (IBAC, 2016) illustrates a number of well-regarded corporations and consultancies being

involved in the project – including Oracle, KPMG, Deloitte, PWC, Boston Consulting Group, and Ernst & Young. While this study recognises a complex set of causes in the failure of the Ultranet, the number of consultancies involved in the project suggests an excessive focus on strategic and goal-oriented action. This study proffers the suggestion that an excessive focus on market-based actors focusing on strategic action as part explanation for the failure of the Ultranet.

The failure of the Ultranet project challenges the assumption that strategic and instrumental solutions offered by markets can provide the kind of system integration the project sought to achieve. The Ultranet sought to integrate activities related to teaching, learning, curriculum, and assessment to improve communication across stakeholders including teachers, school leaders and the bureaucracy (P. Griffin & Woods, 2006; Pearson, 2012). In terms of the conceptualisation of social integration and system integration described in Chapter 5 (Figure 5), the stated goals of the Ultranet project suggest that it was designed to provide system integration in support of processes of social integration provided by schools. This is consistent with Habermas (1975/2005) who describes increased imperatives for system integration as societies advance along technological and social dimensions. Habermas (1983/1996) also suggests that this requires communicative action and postconventional reasoning, and not instrumental action as proffered by markets.

One possible explanation for the failure of the Ultranet is a lack of communicative action to establish common understanding among a plurality of stakeholders. This claim is supported by the government auditor for the project, who identified three business cases, where

None of its three business cases had a well thought out needs analysis or gave comprehensive options to deliver the project. (Pearson, 2012, p. 20)

Only the third business case was accepted by government, and the auditor goes on to argue that there was not a good case for accepting it. These observations suggest a lack of documented communicative action in the project.

The failure of the Ultranet project supports the hypothesis that while markets can relieve the demands on communicative action, markets also have difficulty providing

postconventional reasoning. Market-based private corporations and consultancies are inherently strategic and goal-oriented, which is somewhat incompatible with the kind of ideal and impartial role taking (Habermas, 1983/1996) required for system integration in advanced plural societies. While not precluding the possibility of markets providing this sort of service, a stronger integrative system responsible for developing sound overarching principles might assist projects such as the Ultranet to meet the needs of society and its citizens.

Markets and unhelpful dichotomies

As described in Chapter 1, one of the motivating issues for this study is a malaise in educational assessment. Masters (2013, p. 2) associates this malaise with the rise of dichotomies in educational assessment; such as quantitative versus qualitative, and formative versus summative. This study identifies that the unhelpful dichotomies observed by Masters (2013) are promoted through markets.

The unhelpful dichotomies observed by Masters (2013) can, in part, be attributed to differentiated marketing strategies in education as described by Kotler and Levy (1969). Differentiated marketing requires products to be differentiated to appeal to certain markets, with messages differentiated to appeal to certain audiences. In this way, the strategic action of marketing relieves the burden on communicative action and postconventional reasoning for audiences in target markets. That is, actors in target markets can engage with complex ideas and complex technological system without engaging with broader concepts and principles. Markets, for example, can be differentiated based on those making intended interpretations in central agencies and those making actual interpretations in local contexts. In this way, differentiated marketing potentially fragments educational assessment when no coherent synthesis across dichotomies is developed to enable coherent understanding across intended and actual interpretation (Moss, 2016).

The unhelpful dichotomies arising from commercial vendors can be exacerbated by, what Ellis and McNicholl (2015, p. 67) describe as, *academic capitalism* in the field of education (Jessop, 2017; McClure, 2016). Academic capitalism pursues market and market-like activities that, in the context of market differentiation, amplifies and facilitates the dichotomies described by Masters (2013) in a similar way to commercial

markets. That is, where academic research has the potential to provide for coherence of meaning across intended and actual interpretation, the phenomena of academic capitalism and differentiated marketing works against such a synthesis.

An example of how a desire to differentiate can diminish the communicative capacity of educational discourse is illustrated by the concept *Assessment for Learning*. This relates to one of the dichotomies Masters (2013) describes as fracturing the field of educational assessment. *Assessment for Learning* emerged from the Assessment Reform Group (ARG, 1999, pp. 6-7) who sought to differentiate *Assessment for Learning* from other forms of assessment, particularly formative assessment, by placing greater emphasis on classroom assessment as a way of promoting learning. The ARG (1999) present sound conceptual reasons for distinguishing *Assessment for Learning* from formative assessment, however the two forms are nevertheless often treated synonymously. Harlen (2009), for example refers to the two synonymously as “(formative assessment, or AfL)” (p. 247). This differentiation is in the context where both Masters (2013) and Newton (2007) consider the distinction between formative and summative unhelpful. In the midst of this, Masters (2013) seeks to make another distinction through a unifying principle. While it is difficult to identify who can benefit from this kind of market differentiation and how they might benefit, from a teacher, school, or policy-maker perspective, this kind of differentiation can only be confusing.

Closing remarks

This chapter explored education’s relationship to the economy and polity through a schema developed by Parsons and Smelser (1956/2005). It has addressed how the rise of the knowledge-based economy (Drucker, 1968/1992), new divisions of labour driven by technological development (Levy & Murnane, 2004), and broader social development (Fraser, 2013; Rodwell, 2016), foregrounds education. In this foregrounding, education is considered to have stronger links to both the economy and polity than traditional framings.

In exploring the relationship between the economy and polity. These tensions have been framed here as being managed by an integrative system as described by Parsons and Smelser (1956/2005). A number of tensions have been identified that will be further explored in Chapter 8, along with issues pertaining to the integrative system.

This chapter has also engaged neoliberal thinkers such as Friedman (1962/2002) and Popper (1945/2002) who base their arguments on notions of “approved” in terms of educational services, without detailing procedures or principles for this approval. In failing to address the nature of approval through community consensus, this study finds neoliberalism inherently conservative by reifying existing practices and failing to address progress narratives in society. This contrasts to how Habermas (1992/1998) approaches these matters through deliberative democracy.

This chapter has characterised marketing (Kotler, 1972; Kotler & Levy, 1969) as a form of social action consistent with Habermas (1981/1985), and as having the capacity to relieve the burden on moral-practical reasoning in coordinating society. However, while markets can assist in addressing complexity, markets also change the nature of social integration. Through markets, social integration no longer becomes a universal concern but a concern among those with shared interests and the resources to engage in those interests. In this way, the markets of neoliberalism are distinct from pragmatic approaches to social integration, a tension further explored in Chapter 8

This study has identified class, sexuality, gender, and race as potential variables through which education markets can be segmented. While this is a poignant argument, this study, does not pursue this argument. Instead, this chapter has identified issues arising out of markets in educational assessment. These issues relate to a proliferation of purposive commercial educational assessment programs, particularly those pertaining to the transition from secondary to tertiary education. Markets can also adversely affect educational assessment design and implementation, particularly for projects requiring integration along technological and social dimensions and which require the generation of common understandings. This chapter has found that marketing, including markets of academic capitalism (Ellis & McNicholl, 2015), might explain the proliferation of unhelpful dichotomies in educational assessment observed by Masters (2013).

The next chapter will build on the framing developed in this chapter. Where this chapter addressed educational assessment’s inherent relationship to the economy and polity, the next chapter identifies educational assessment validity as simultaneously addressing ethics and science.

Chapter 7. Colonisation of educational assessment by science

This chapter addresses tensions between science and ethics in educational assessment. The first section frames this tension through the work of Borsboom et al. (2003, 2004), who argue that for an assessment to be valid its latent variable needs to exist, exist independently of measurement, and be causally responsible for measured outcomes. It is through engaging with this argument that support emerges for Baird et al. (2017) who make a case that educational assessment be distinguished from psychological assessment. Borsboom et al. (2003, 2004) usefully describe a realist stance, and the second section addresses contemporary ethical issues that emerge from the realist stance, such as the claim that educational outcomes are caused by physical attributes such as DNA.

The third section provides a formal explication of performativity and its association with scientific reasoning as elaborated by Lyotard (1979/1984). Performativity is contrasted to the performative attitude elaborated by Habermas (1981/1985) as engaging with the cognitive, social and aesthetic dimensions. The social and aesthetic dimensions are identified as emerging areas of importance for educational assessment.

The fourth and final section addresses education system management and explores how the realist stance facilitates scientific approaches to system management including strategic management. The section closes with a postmodern rejoinder that draws on Baudrillard (1981/1994) and Luhmann (1984/1995). The rejoinder alerts to the possibility that educational assessment can evolve to no longer reflect the reality of educational practice. This is to foreground a concern for ensuring that symbolic media such as educational assessment continues to reflect activity in society so that it meets the needs of citizens.

7.1. Educational assessment and psychological assessment

Chapter 6 illustrated how education, and educational assessment, addresses a student's relationships to the economy and polity. In that framing, educational assessment simultaneously addresses properties of autonomous students and properties of society independent of the student. How educational assessment approaches properties attributable to society and properties attributable to students pertains to contemporary debates in assessment validity.

Messick (1989) considers social consequences an essential aspect of validity, an aspect that addresses “the meaning, relevance, utility, and value implications” (p. 84) of test design and use. This is consistent with the case that educational assessment coordinates activity in society in ways that address ethical and scientific issues simultaneously. The position of Messick (1989) is consistent with Habermas (1981/1985) who argues that the objective world of science, and the social world of normative ethics, are simultaneously raised in all validity claims. Following Habermas (1981/1985), ethical validity is based on subjective feelings where validity is redeemed intersubjectively through moral-practical rationality. Similarly, scientific validity is based on subjective perceptions that are redeemed intersubjectively through cognitive-instrumental rationality (Habermas, 1983/1996; Toulmin, 1950/1970). This chapter describes how both can, and are, treated simultaneously in contemporary educational assessment practices and in a way that suggests that the position of Messick (1989) is consistent with that of Habermas (1981/1985).

Validity in assessment

One factor identified in this study that inhibits educational assessment from fully addressing ethics is that it is currently linked to psychological assessment which is more amenable to a scientific focus. *The Standards* (AERA/APA/NCME, 2014), for example, has separate chapters for psychological testing, workplace testing, educational testing, and program evaluation, but has one common chapter on validity that applies to all forms of testing. As will be elaborated through this chapter, contemporary educational assessment recognises the importance of sociocultural and sociopolitical influences in its practices which simultaneously address ethics and science (Crocker, 1997a; Messick, 1989; Shepard, 2016). This study also notes that Borsboom et al. (2003, 2004) make a particularly coherent explicit argument against this position. In light of the compelling arguments made by Borsboom et al. (2003, 2004), this study works towards the elaboration of two validity paradigms clustered around notions of educational assessment and psychological assessment.

In making the case that educational assessment be considered distinct from psychological assessment is not to suggest that there is no overlap between the two. Both education and psychology share concerns. Instead, contemporary practices in educational assessment

simultaneously and indivisibly address properties of the student that are intersubjectively related to properties in their sociocultural environment. In this sense, the properties of a student in education can never be isolated from the sociocultural environment in which they live. Further, the sociocultural environment is constructed by humans through cultural practices that evolve in response to technological development and social progress. This study identifies branches of psychology that share similar concerns to education in combining personal traits with cultural and social factors, but that psychology also has branches that focus on isolating personal traits.

The case for two paradigms emerges from contemporary validity debates, where Borsboom and Wijsen (2016) argue in favour of a “distinction between facts and moral values” (p. 282), Shepard (2016) argues that

To act as if value choices and ethical decisions are outside of science is to ignore the value-laden nature of the scientific process ... (p. 276)

It is argued here, that the two contrasting positions are best resolved through differentiating two forms of assessment with two distinct validity paradigms.

Realism and constructivism – ontologies for the latent variables

The distinction between realism and constructivism, as described by Borsboom et al. (2003), provides a point of entry into differences between educational assessment and psychological assessment. Borsboom et al. (2003) argue in favour of realism and argue from the perspective of psychological assessment. This claim is based on references that Borsboom et al. (2003) make to attributes such as extraversion, personality, spatial ability, self-efficacy and attitudes in their argument, these are generally associated with psychological assessment. Borsboom et al. (2003) do not refer to educational domains such as mathematics or reading in making their case to suggest a distinct psychological orientation.

Borsboom et al. (2004) make the realist case through the claim that

a test is valid for measuring an attribute if and only if (a) the attribute exists and (b) variations in the attribute causally produce variations in the outcomes of the measurement procedure. (p. 1061)

Further, Borsboom et al. (2003) associate realism with the requirement

to have a match between theoretical relations and relations in reality, these relations in reality have to exist quite independently of what we say about them. (p. 209)

Of significance is the importance that the realist stance places on measures being independent of what is said about the latent variable. This implies a decoupling of measures from the linguistic explication of the latent variable, or a decoupling of the numerical component from the linguistic component of educational assessment.

By way of contrast, Borsboom et al. (2003) go on to describe constructivism as considering latent variables as human constructions, and operationalism as a numerical trick that simplifies observations. Constructivism does not make similar claims to truth in the way that realism does towards the correspondence between theory and reality. Instead, constructivism considers “that people construct this relation themselves” (p. 211). This characterisation of constructivism by Borsboom et al. (2003) is consistent with a pragmatic approach towards educational assessment. This is to make the point that there are no definitional issues identified by this study over how Borsboom et al. (2003) describe and characterise realism, constructivism and operationalism.

Three features of realism identified by Borsboom et al. (2003, 2004) are pertinent to the present discussion. First, Borsboom et al. (2003, 2004) emphasise causality, where variations in observed outcomes, such as student responses to a test, should be able to be attributed to variations in a latent variable that exists in reality. That is, for a test to be valid an identifiable entity responsible for a student’s responses must exist within the student, and in a way where the entity is shared across individuals. This emphasis on causality attributed to something real, and not something socially constructed, is open to empirical approaches that seek to find cause through categorical variables such as DNA, or attributes of sex, race, skin colour, sexuality or similar.

An interest in categorical explanations relates to the second feature of the argument made by Borsboom et al. (2003, 2004), which has a focus on intra-individual or within subject structures, and inter-individual or between subject variation. That is, realism has a focus on structures that are properties of the student themselves, and not properties of a

student's socially constructed reality including the economy or polity. Realism's focus on the student in isolation of their sociocultural environment conflicts with approaches to social integration and socialisation described in Chapter 5.

The third feature is that realism considers the latent variable to exist independently of what is said about it, and that it exists independent of measurement. When this validity requirement is applied to the field of educational assessment, then the validity of an assessment would not rely on its correspondence to, for example, a curriculum statement, a domain explication or an assessment framework. Instead, the latent variable would refer to an attribute similarly self-evident to two researchers who sought to measure the same trait. It is this feature of realism that pertains most directly to reasons identified by Habermas (1981/1985) for developing the concept of communicative action and is explored next.

Latent variables without explication

Chapter 5 described how Habermas (1981/1985) developed the theory of communicative action as he found the conceptualisation of action by Weber (1915/1964) inadequate as it presupposes a cultural context and common understanding. Through its requirement that an attribute exist independently of explication and what is said about it, realism also presupposes a self-evident cultural definition of latent variables in education such as mathematical or reading ability. In a similar way to how Habermas (1981/1985) considers the presuppositions of Weber (1915/1964) inadequate, this study finds realism in educational assessment inadequate as it presupposes domain and construct definitions as well as rendering these definitions silent.

The realist stance lends itself to assumptions that self-evident traits exist as real. For example, one teacher's question of another on how many boys and girls they have in their class is often considered unproblematic. That is, the category of boy and girl is considered generally unproblematic and sufficiently "real" for such a question not to require communicative action regarding the propositional content of the question. However, even on this simple question, Gates (2011) estimates that 0.3% of adults would find the question problematic on the basis of transgender. That is, even this superficially simple question can exclude a small but not insignificant amount of the population in cases where there is no broader common understanding between the teachers around transgender. This

is to provide clear case support that even on the simple categorical variable of sex and gender the realist assumption can be problematic.

The realist stance becomes more problematic in case of curriculum and gender. For example, in a paper on the school curriculum for science, Fensham (2016) reports that

The research suggested that girls prefer to learn concepts in their social context rather than abstractly and that they prefer to respond to structured and extended response questions rather than multiple choice ones. (p. 172)

There are two socially constructed concepts implied in this statement. First is the notion of “girl”, while largely unproblematic, it may not be universally accepted (Gates, 2011). The second is how the somewhat socially constructed variable of “girl” relates to the socially constructed latent variable of the educational assessment. Where the constructivist stance of Fensham (2016) might seek recourse to substantial argument (Kane, 2006; Toulmin, 1958/2003) over assessment frameworks, the realist stance considers these questions unproblematic and self-evident to render them silent.

The complex and humanly constructed nature of gender and sex is fraught and is an anathema to the realist stance (Butler, 1990/2007). Even the simple claim by Fensham (2016) that girls prefer to learn differently could be charged with *essentialism* in “theorizing gender as both an identity and a mark of difference” (Diquinzio, 1993, p. 1). While the language used by Fensham (2016) is appropriately soft, it alerts to the need for the paralogical reasoning and little narratives advocated by Lyotard (1979/1984) to insinuate the marginalised into the universal. Fensham’s (2016) argument is paralogical and fraught as it involves identifying a socially constructed category as being different so that that people assigned to that category can be treated as the same.

As discussed in Chapter 2, differences between the realism and constructivism is not so much about ontology in educational assessment, but more a question of how educational discourse is framed through either a realist or constructivist stance. The realist stance, for example, is consistent with the wants as fixed assumption made by Friedman (1962/2008) in monetarist economics when applied to curriculum and definitions of the latent variable as being similarly fixed. The realist stance is consistent with Friedman’s (1962/2002, p. 89) use of terms such as “approved educational services”, and “approved school” as being

largely unproblematic and self-evident. The realist stance is also consistent with certain conservative ideological positions such as that of Caldwell and Hayward (1998, p. 12), described in Chapter 1, who consider feminism and a sensitivity towards the aspirations of minorities as challenges to authority. That is, the realist stance is consistent with not having to discuss or consider the construction of categories such as gender (Butler, 1988, 1990/2007), race (Gandhi, 1998; Said, 1978/1994; Spivak, 1985/2010), or curriculum, in a way that might challenge what McIntosh (1986) describes *white male privilege*.

The realist stance is antagonistic towards the dialectic inherent in the differential performance of girls and boys in assessments, an issue explored by Wiliam (2010) in a discussion on constructs. Wiliam (2010) identifies that differences were once attributed to factors such as biological predisposition and genetic factors, to conclude that

in recent years, the extraordinary decline in the size of sex differences in mathematics performance has provided strong evidence that observed sex differences are primarily of environmental, rather than genetic, origin. (p. 274)

Where the realist stance in advocating for causality might seek to attribute variation in test scores to causes such as genetics and gender, the constructivist stance might seek to interrogate how gender is reflected and constructed in the latent variable (Butler, 1990/2007). This is consistent with Wiliam (2010), who argues that matters of equity are best conceptualised “as issues of construct definition and construct choice” (p. 268). This view is consistent with Fensham’s (2016) approach to gender, to suggest a dialectic relationship between how a latent variable is constructed, and how identities such as gender are constructed. The realist stance is unresponsive to arguments around these social constructions.

This study does not seek to arbitrate on what are appropriate categories in terms of sex and gender, or on what is an appropriate construct choice or construct definition. Instead, this study seeks to expose these matters as socially constructed and open to argument. The realist stance lends itself to analytical arguments as defined by Toulmin (1958/2003), where analytical arguments do not question predicates such as boy, girl, or curriculum content. The constructivist stance seeks to open predicate definitions through Toulmin’s (1950/1970, p. 74; 1958/2003, p. 63) substantial argument that addresses what is

beautiful, what is good, what is right, and what is valued. Where pragmatism seeks to open these issues to the better argument, the realist stance renders these arguments silent.

The realist stance advocated by Borsboom et al. (2003) may therefore not be suited to educational assessment but may be appropriate for some forms of psychological assessment explored next.

Six perspectives on psychology

Fernald (2007) explores six perspectives towards psychology from which the *biological* and *behaviourist* (Pavlov, 1927; Skinner, 1938, 1957) perspectives best align with the realist stance of Borsboom et al. (2003). The realist stance is consistent with the biological perspective that explores physical entities such as the organisation of the brain, the nervous system, the cerebral cortex, synapses, neurons and cells. These entities are real to the extent that they can be physically observed and empirically studied. Similarly, the behaviourist perspective aligns with the realist stance through a similar ontology which considers “biological and other physical phenomena as the proper and eventually adequate explanation of mental activities” (Fernald, 2007, p. 169). This is to provide clear case support to the argument that some forms of psychological assessment align with a realist stance.

The biological and behaviourist perspectives also lend themselves to inter-species reproducibility, with early proponents such as Skinner (1938, 1957) and Pavlov (1927) both conducting experiments on animals (Fernald, 2007, pp. 130-168). This supports the claim that behaviourism, like the biological approach, is not tied to specific social contexts, or even to a species.

There is some overlap between behaviourism and education however. Therapies within the behaviourist tradition include positive and negative reinforcement (Fernald, 2007, pp. 130-168). A similar focus on positive and negative reinforcement is found in some pedagogical approaches such as direct methods of instruction. For example, Engelmann et al. (1988) describe direct instruction as involving “scripted presentation of lessons, small-group instruction, reinforcement, corrections, and procedures to teach every child” (pp. 305-306). This is to suggest some overlap between psychology and education in terms of behaviourism. However, direct instruction does not represent the full range of

approaches to pedagogy, with constructivist approaches also used in education (Dewey, 1916/2011; Larochelle, Bednarz, & Garrison, 1998; D. C. Phillips, 1995).

When psychology is considered from the six perspectives described by Fernald (2007), some overlap emerges between the fields of psychology and education. The psychoanalytic, humanistic, and evolutionary perspectives share with educational assessment a concern for the social and cultural environment. Cognitive psychology is another perspective described by Fernald (2007), a perspective that directly relates to educational assessment with its concern for “processes and the patterns ways (sic) through which people think and act” (Mislevy, 2006, p. 296). Snow and Lohman (1989), as well as Mislevy (2006), describe cognitive psychology in terms not so much concerned with scores related to latent variables, but more about processes used by individuals towards processing information in the environment. This knowledge can then be used to feed into educational assessment design. This is to suggest a borderline case between educational assessment and psychological assessment. While there is some overlap between educational and psychological perspective towards assessment, the biological and behaviourist perspectives point to a clear case difference.

Furthermore, as described in Chapter 2, the distinction that Reckase (2017) makes between the continuum and domain sampling models also points to differences between psychological and educational assessment. Reckase (2017) describes psychological assessment as concerned with estimating a student location on a continuum, and educational assessment concerned with estimating what proportion of an explicated domain a student has acquired. The unifying principle proposed by Masters (2013) also becomes problematic on this point through its focus on locating students on a scale. In terms described by Reckase (2017), Masters’ (2013) principle is better suited to psychological assessment that seeks to locate its subject on a continuum. In associating educational assessment with a domain sampling model, Reckase (2017) provides further support to the case that psychological and educational assessment proceed in separate paradigms (Baird et al., 2017).

Different approaches to ethics and science

Psychological assessment and educational assessment lend themselves to different approaches towards science and ethics. The separation between facts and moral values

argued for by Borsboom and Wijsen (2016) is impossible to sustain in educational assessment given its sociocultural nature. Borsboom and Wijsen (2016), correctly argue that science and politics have different goals, and different interests,

science is interested in questions about truth and falsity, politics is concerned with what is right or wrong. (p. 282)

This description is consistent with the position of Habermas (1981/1985) in relation to science and ethics. However, as described in Chapter 5, Habermas (1981/1985) considers that a relation to both science and ethics, as well as the aesthetic, are addressed simultaneously in communication in the lifeworld. It is to the question of whether science and ethics can effectively be separated in educational assessment that has generated debate in the field of validity in recent decades (Crocker, 1997a; Messick, 1989; Newton & Baird, 2016a; Shepard, 2016).

While it may not be possible to separate science and ethics, or truth and ethics, in contemporary educational assessment, it is possible to separate them into two steps in certain forms of psychological assessment. For example, an assessment might include randomised controlled trials such as those used in health care research (K. F. Schulz, Altman, Moher, & Group, 2010). In such a trial, ethical considerations might be addressed in a first step using guidelines such as the *Declaration of Helsinki* (World Medical Association, 2013). In this case, ethics could be formally considered separately prior to the research being conducted. After research has been approved on ethical grounds, it can proceed accordingly through scientific and objective principles. This study finds educational assessment as not having a similar capacity to isolate ethical and scientific considerations. At a minimum, and as will be shown later in this chapter, educational assessment builds on a pattern of a teacher asking a student a question, and it is impractical for a teacher to seek ethical approval before asking each question. This becomes particularly difficult when teachers engage students in discourses of application, including the provision of feedback on a student's performance educational assessment (Delić & Bećirović, 2016; Habermas, 1994; Hattie & Timperley, 2007; Lipnevich et al., 2016; Szypszak, 2015).

The structure of the PISA also illustrates the point that science and ethics, or science and politics, are difficult if not impossible to separate in educational assessment. The PISA

has a politically oriented governing board that sets its priorities (OECD, 2017e, p. 26), and which interacts with a scientifically oriented technical advisory group (OECD, 2017e, pp. 114,266,268, 460). This relationship is further explored in Chapter 8.

Truth or generalised communication

The argument by Borsboom and Wijsen (2016, p. 281) that assessment validity is concerned with a matter of scientific truth, and the favouring of realism by Borsboom et al. (2003, 2004), is hard to sustain in educational assessment. Validity arguments based on the assumption that the latent variable is real requires the constructs of educational assessment to be stable entities. However, Wu et al. (2017), for example, consider that

the notion of a construct in psycho-social measurements may be somewhat fluid in that definitions are shaped depending on the contexts and purposes of the measurements. (p. 20)

This is to suggest that a construct is not sufficiently stable as an entity to attribute causality. As Wu (2009) shows in comparing the PISA and the TIMSS, how a construct is defined, or how a construct is manifested in the assessment material (Wiliam, 2010), influences the measures obtained in an assessment. This sentiment is reflected too by McCurry (2017, p. 4), as described in Chapter 2, that the type of construct influences the nature of test items. The argument that educational assessment addresses an objective scientific truth, and a property solely pertaining to the student, is therefore difficult to sustain.

The fluidity of constructs described by Wu (2009) suggests that operationalism might be a better term to describe educational assessment. Borsboom et al. (2003) describe operationalism as nothing more than a numerical trick to simplify observations. McCurry's (2017) account around the development of constructs supports the suggestion that constructs are an evolving entity. Wiliam (2010, p. 259) also describes constructs as becoming manifest in the assessment content, to suggest that a construct is only implied through the test content. This suggests that even a claim to constructivism might be an overclaim for educational assessment, and that operationalism might better describe it as a field. Operationalism, like constructivism, is also consistent with Habermas' (1981/1992) concept of media as generalised communication.

This study finds the realist stance as argued for by Borsboom et al. (2003, 2004) as unsustainable in educational assessment. The next section addresses how the realist stance leads to ethical issues in educational discourse.

7.2. The realist stance and ethics in educational assessment

Borsboom and Wijsen (2016, p. 282) argue that facts should be treated distinctly from moral values, and science distinctly from politics, in assessment. For Borsboom and Wijsen (2016), the empirical facts of science should be treated distinctly from ethical considerations and regard for social consequences, an approach that contrasts to that of Messick (1989).

The realist stance towards educational assessment and educational achievement is reflected in research exploring how student test performance relates to attributes such as genetics. How genetics interact with the social environment and social processes in education has been of ongoing concern in education. This relationship is investigated by Kovas, Tikhomirova, Selita, Tosto, and Malykh (2016), who argue that educational approaches based on genetic profiles can be beneficial. Baron-Cohen et al. (2014), on the basis that mathematics ability is heritable, identify a number of genetic markers that are likely to be causally related to mathematics ability. Kautz et al. (2014, p. 8), in a working paper for the OECD, investigate personal attributes that might contribute to work performance and suggest genetics as one factor. This scientific interest in how genetics influences student performance is consistent with the realist stance advocated by Borsboom et al. (2003, 2004).

This study does not seek to arbitrate on how, and to what extent, biological factors and heredity affects student performance. Any effect that these factors have on performance are indeed a matter of science. However, this study does identify that how these factors are addressed in contemporary educational assessment is such that matters of ethics cannot be separated from matters of science. This continues to be explored throughout this chapter. How consideration of scientifically real factors such as biological factors can affect students as citizens is explored next.

Biological factors and differential expectations

One of the central issues that arises when group differences are identified on categories such as sex, race or class is that this can lead to differential treatment and differential expectations for students based on these categories. This pertains to critiques by poststructuralists addressed throughout this study (Butler, 1990/2007; Said, 1978/1994; Spivak, 1985/2010), where discrimination or exclusion based on sex, race and class is justified on scientific grounds (Hunter-Doniger, 2017; Lowe, 1980). As previously discussed, differential outcomes with respect to gender and sex are addressed by Fensham (2016) from a constructivist perspective. Differential expectation from a realist perspective is addressed next through the contemporary work of Didau (2016).

Didau (2016, pp. 402-408), in making an argument that standardised *intelligence quotient* (IQ) tests are not biased with respect to sex and race, argues that test score differences between races and sexes are “largely the products of ... pupil’s IQs and those IQs are, in turn, largely due to variations in DNA” (p. 405). In support of this claim, Didau (2016) cites broader research from Shakeshaft et al. (2013) that claims student performance on educational achievement tests is influenced by genetics. Following this argument, Didau (2016) goes on to suggest that schools

should not be judged according to the extent to which they are closing various achievement gaps, ... [it] is not a coherent prior belief to expect that, in a meritocratic society, the various social classes will not differ somewhat in their mean IQs and innate capability for academic achievement. A powerful body of literature implies that schools will not be able to produce equitable outcomes... (p. 406)

While Didau’s (2016) book is not an academic book, these views are of interest as it illustrates an enduring tension related to *eugenics*, a perspective that seeks to develop or maintain a particular social order through psychometrically supported biological arguments (Hunter-Doniger, 2017; Lowe, 1980). Didau’s (2016) book is also of interest because it is endorsed through a favourable foreword by Dylan Wiliam, an academic expert in the field of educational assessment (Black & Wiliam, 1998; Wiliam, 2008, 2010, 2016).

Didau's (2016) position suggests a realist stance towards the latent variable of educational assessment. That is, Didau (2016) is arguing that outcomes on educational tests are causally linked to a real entity, in this case DNA. Further, Didau (2016) is arguing that a direct social consequence of this empirical evidence is that schools should not be expected to address differences in outcomes between social classes. This study considers this a matter of ethics, an issue that points to a broader tension between empiricism and rationalism in philosophy, a tension earlier explored in Chapter 3, and expanded upon next.

Empiricisms and rationalism, and the is-ought

Didau's (2016) argument alludes to an enduring tension between empiricism and rationalism in philosophy which extends to ongoing debates in educational assessment. Messick (1989), for example, considers validity the degree to which empirical evidence and theoretical rationales support inferences and actions arising from test scores. Cizek (2012a, p. 35; 2016, p. 214) considers Messick's definition central to contemporary notions of educational assessment validity, while claiming that no synthesis of evidence as described by Messick has ever been produced. This suggests that tensions between empiricism and rationalism in the field of educational assessment are ongoing, with Didau (2016) presenting a contemporary view.

The argument presented by Didau (2016) is often approached in terms of the *is-ought* problem first described by Hume (1739/1985). It is a tension explored by Wiliam (2010) and Biesta (2009, 2010) in the context of educational assessment and discussed later in this chapter. Hume (1739/1985, p. 469) challenges the notion that an *ought* follows from an *is*, and challenges how explanations follow from observations, to conclude that the relationship between an observation and a deduction arising from it as inconceivable. The relationship between *is* and *ought* is central to both intended interpretation and actual interpretation. The relationship pertains to questions on what intended interpretations *ought* to be reported, and what action *ought* to arise from an actual interpretation.

Didau (2016) is following the tradition of Hume (1739/1985) in taking a realist stance that focuses on DNA in the interpretation of educational assessment data. Garrett (2015, p. 3) describes Hume (1739/1985) as being primarily concerned with empiricism that emphasises observations as dictating the content of theorising without invoking other

entities, ultimate principles, or basic values. This empirical focus is reflected in Didau's (2016) position that causally links the empirical observations of class, gender and race with the empirical observations of student outcomes without recourse to other principles.

The position of Kant (1772/1977) contrasts to that of Hume (1739/1985). Kant (1772/1977, pp. 2-3) was motivated by Hume's (1739/1985) work on cause and effect to take a contrasting view on the nature of the *a priori* and ultimate principles. Hume (1739/1985) considers humans to be ignorant of any "ultimate principle" which bind cause and effect together, and that these principles are "an illusion of the imagination" (p. 267). In contrast, Kant (1781/2016, p. 137) considers that experience or empiricism never leads to true or strict judgements, and that for a judgement to be considered universal, it needs to be considered through an *a priori*. As described by Schneewind (2002), "Kant rejected Hume's naturalism and insisted that we [humans] are not merely natural beings, like the animals" (p. 87). Where Hume (1739/1985) emphasises the empirical in truth, Kant (1781/2016) favours that judgements be tempered by rational consideration. On this matter, the position of Habermas (1983/1996) is closer to that of Kant (1781/2016) by requiring that universal norms be acceptable to all those affected through principles of discourse.

The position taken by Didau (2016) therefore illustrates an ongoing tension in the field of educational assessment. A tension that can be characterised as one between Kant (1781/2016) and Hume (1739/1985), between empiricism and rationalism, or as an is-ought dilemma. It also relates to the division between science and moral values, or science and ethics, as described by Borsboom and Wijsen (2016). How contemporary practices in educational assessment address issues such as those emerging from Didau's (2016) observation is explored next.

Interpretation in practice

How data is interpreted and what consequential decisions arise is an ongoing concern in educational assessment. Wiliam (2010) draws on the is-ought argument of Hume (1739/1985) in making arguments on how to interpret educational assessment data. For example, Wiliam (2010) draws on data from a number of exams that have good predictive properties for further education and have differential outcomes for students of colour. Wiliam (2010) considers that these exams predict well and differentiate on colour as an

is, and that no *ought* consequences arise from this *is*, such as lowering admission requirement for people of colour.

Biesta (2009) also draws on the is-ought of Hume (1739/1985) to take a different stance to Wiliam (2010). Biesta (2009) argues

that if we wish to say something about the direction of education we always need to complement factual information with views about what is desirable. (p. 35)

Biesta (2009) is here following in the tradition of Kant (1781/2016), where empirical educational assessment data is considered on the basis of some principle or value. Further, this principle or value is a matter of ethical reasoning to support the argument that ethics and science are simultaneously combined in the design and use of educational assessment.

Civil Rights Act of 1964, United States

The Standards (AERA/APA/NCME, 2014) encourage a more circumspect and cautionary approach when compared to that of Didau (2016). *The Standards* (AERA/APA/NCME, 2014) stipulate that

When aggregate scores are publicly reported for relevant subgroups—for example, males and females, individuals of differing socioeconomic status, individuals differing by race/ethnicity, individuals with different sexual orientations, individuals with diverse linguistic and cultural backgrounds, individuals with disabilities, young children or older adults — test users are responsible ... for including cautionary statements ... (p. 71)

It is Didau's (2016) failure to provide any cautionary statement that contrasts to the approach of *the Standards* (AERA/APA/NCME, 2014). Throughout *the Standards* (AERA/APA/NCME, 2014) reference is made to attributes such as race, ethnicity, and sex. There are also extensive references to legal requirements that largely arise from the *Civil Rights Act of 1964* in the United States where test bias with respect to race is associated with long standing and ongoing political and legal developments (Cole & Moss, 1989; Gordon Commission, 2013, p. 27; S. E. Phillips & Camara, 2006). S. E. Phillips and Camara (2006) describe several legal cases addressing race and test bias

subsequent to the introduction of the *Civil Rights Act of 1964* in the United States. The cautionary approach to assessment detailed in *the Standards* (AERA/APA/NCME, 2014) arises from this tradition of legal contestation on matters of race and minority groups.

The Standards (AERA/APA/NCME, 2014) is a precise document informed through several iterations that incorporate ongoing developments in science as well as legal precedents. Several legal decisions and precedents are described by S. E. Phillips and Camara (2006). *The Standards* (AERA/APA/NCME, 2014) attempt to balance both scientific and ethical concerns, and empirical and rational concerns, in a justifiable way in light of these legal precedents. Some of these considerations are addressed now.

The empirical argument and prediction

The focus on empiricism in educational assessment is often driven by a concern for prediction. How well educational assessment data predicts future behaviour, or correlates with current behaviour, has been an ongoing concern in assessment validation (Cronbach & Meehl, 1955; Kane, 2006; Messick, 1989; Newton & Shaw, 2014). For example, how well results reported through secondary school certificates predict performance in higher education is of particular contemporary interest (Anderton, Evans, & Chivers, 2016; B. Griffin et al., 2018; Mercer et al., 2018; Poole et al., 2012).

In terms of prediction and differential performance by subgroups, Wiliam (2010) argues that issues of bias are best addressed through construct choice rather than validity. This exposes deeper arguments. Arguments on how constructs are defined are not evident in the surface structures of educational assessment, and as described in Chapter 3, are addressed in Tier-2 of the three tiers of educational assessment. Tier-2 involves substantial arguments regarding design (Toulmin, 1958/2003). In discussion on the SAT¹ college admission test in the United States Wiliam (2010) argues that differential performance is a matter of construct choice and that

blaming the SAT for the failure of students of color to gain admission to, and to thrive in, the most selective colleges is unlikely to do anything to improve the situation. The SAT works as well as it does because it is exquisitely tuned to the system in which it operates. (p. 269)

¹ SAT is a redundant acronym for a college admission test in the United States - www.collegeboard.org

Wiliam (2010) argues that differential performance results from both construct choice and also that minority groups do less well on the SAT “because they are less prepared for college” (p. 268). Wiliam (2010) here alludes to two arguments that are not empirical or analytical arguments, but substantial ethical arguments related to the construct being assessed and opportunity to learn. These two issues are addressed next.

Construct choice and domains in society

How constructs are conceptualised can affect test performance of certain subgroups of the population. As previously discussed, Fensham (2016) describes changes to assessment and curriculum in science during the 1980s in Victoria, where the imperative for change was a “concern about the lack of gender equality in school science” (p. 172). This led to developments that changed how subjects like physics were taught and assessed. Changes included what type of content was covered. Fensham (2016) describes how a construct is conceptualised, and how it is assessed through educational assessment, affects performance by subgroups of populations such as girls.

While changes to constructs assessed in educational assessment might be able to ameliorate differential performance, there is nevertheless a countervailing pressure from outside of education. S. E. Phillips and Camara (2006) describe challenges arising from bias against minority groups in terms of differential impact. S. E. Phillips and Camara (2006) go on to describe that the United States legal system generally accepts the defence that differential impact is justified if a “test is necessary to achieve an important business goal and that the test is job-related” (p. 733). In broader terms used by this study, differential impact on subgroups of population is here justified with respect to the qualification purpose of education discussed in Chapter 4. That is, bias against minority groups can be justified based on how work is conducted in the economy and how an educational assessment relates to this work.

A broader dialectical relationship between education and the economy becomes evident in this context. That is, should the qualification concern of educational assessment reflect what work is carried out in the economy, or should educational assessment have some aspirational component to reflect the way work *ought* to be done in the economy. For example, when Fensham (2016) describes the subjects of physics as responding to gender bias by focusing increasingly on the social world and the usefulness of physics in society,

is this a gender issue, an issue of the construct, or an issue of society? That is, is this change in construct definition an expression of how broader society, that includes women, wishes the field of physics to be conducted. This study does not seek to arbitrate on this matter but seeks to highlight how constructs are part of a dialectic between education and the economy, and a dialectic between education and society more broadly.

The AGIL scheme described in Chapter 6 illustrates how construct definitions relate to ethics. In the AGIL scheme the integrative system is responsible for mediating tensions between how constructs are defined in educational assessment, and how these constructs relate to work in the economy and their legitimacy with respect to the polity. Where neoliberalism would see these tensions addressed through markets that tend to reify constructs in a status quo, Habermas' (1998) pragmatic tradition would see these tensions addressed through deliberative democracy which incorporates ethics.

Opportunity to learn

The opportunity to learn, which Wiliam (2010) provides as one explanation for differential performance on tests, also raises broader ethical and legal issues. *The Standards* (AERA/APA/NCME, 2014) describe opportunity to learn as

the extent to which individuals have had exposure to instruction or knowledge that affords the opportunity to learn content and skills targeted by the test... (p. 56)

S. E. Phillips and Camara (2006) describe opportunity to learn in terms of curricular validity and as the subject of legal cases. These legal cases resulted in the requirement that assessment agencies give notice on what content is going to be assessed on a test. The extended focus given by *the Standards* (AERA/APA/NCME, 2014) towards opportunity to learn and curricular validity further emphasises the ethical focus of educational assessment. That is, educational assessment does not address a real entity that is a property of the student as argued by Borsboom et al. (2003, 2004). Instead, and following *the Standards* (AERA/APA/NCME, 2014), educational assessment is ethically and legally bound to address curriculum that is publicly declared and a human construct (Pinar, Reynolds, Slattery, & Taubman, 2008).

The notion of opportunity to learn directly relates to Habermas' (1981/1985) notion of strategic action described in Chapter 5. As described by S. E. Phillips and Camara (2006), opportunity to learn and curricular validity “requires the state to disseminate information about graduation test requirements to all affected students well in advance of implementation” (p. 735). This supports the notion that educational assessment is an openly strategic social action, where curriculum forms an important part of this social process in declaring what effect the state seeks to achieve in the student. From this, a failure to declare what content is to be assessed through curriculum or test design can lead to, in Habermas' (1981/1985) terms, systematically distorted communication or manipulation.

Fairness

Ethical issues in educational assessment are largely addressed under the rubrics of fairness and due process. S. E. Phillips and Camara (2006) describe fairness as addressing “testing procedures that are arbitrary, capricious, or fundamentally unfair to examinees” (p. 734). Where opportunity to learn is a fundamental aspect of fairness, due process addresses matters pertaining to liberty and can be challenged on both procedural and substantive grounds. Ethical and legal issues arising from fairness are complex and detailed. Fairness and ethics are extensively addressed in educational assessment and are an inextricable part of the field (Brown & Harris, 2016; Camilli, 2006; Masters, 2002; Wesolowski, Wind, & Engelhard, 2015; Zieky, 2016).

Fairness also relates to the provision of *appropriate accommodations* for educational assessment for students with disabilities. This further supports the claim that educational assessment addresses a socially constructed relationship between education and the student (S. E. Phillips & Camara, 2006, p. 249; Thurlow & Kopriva, 2015). That is, a student's biological disposition on ability may require, on ethical grounds, that they be treated differentially in educational assessment. These approaches to fairness provide a compelling case that education, and educational assessment, is a sociocultural practice where values are implicated in interpretation, and where science cannot be separated from ethics.

This study does not seek to arbitrate on Didau's (2016) claim that test scores are causally linked to DNA, nor whether it is a coherent prior belief to expect equality of outcomes.

Instead, this study observes Didau (2016) as being imprudent. In making the claim that it is not a coherent belief to expect all subgroups of a population to achieve equally well, Didau (2016) lacks the caution advised in *the Standards* (AERA/APA/NCME, 2014). It is a claim that also lacks the nuanced ethical approach described in *the Standards* (AERA/APA/NCME, 2014). While Didau's (2016) position is consistent with the realist stance that seeks to separate ethics from science in educational assessment (Borsboom et al., 2003, 2004; Borsboom & Wijsen, 2016), it is inconsistent with contemporary practices in educational assessment.

Paradigm shift

This study identifies a compelling case favouring a paradigm shift in assessment validity and one which follows from the claim by Baird et al. (2017) to distinguish educational assessment from psychological assessment. Current approaches consider educational assessment and psychological assessment in a similar fashion. *The Standards* (AERA/APA/NCME, 2014), for example, do not differentiate validity for educational assessment and psychological assessment. This combined approach is reflected in an extensive explication of the history of validity in assessment by Newton and Shaw (2014). A new paradigm could see educational assessment developing its own approaches to validity where these approaches incorporate notions of legitimacy that address the relationship education has with the polity and economy. Newton and Baird (2016a) not only observe that not much progress has been made since Crocker (1997a) addressed validity, but that “new controversies have arisen, including debate over the relationship between validity and truth” (p. 173) (Borsboom & Markus, 2013; Borsboom et al., 2004). Kuhn (1962/1970, p. 37) describes that scientific communities work within paradigms and choose the problems to be addressed and govern the extent to which they are considered. New paradigms emerge when there are failures in problem-solving and the emergence of crises more generally.

Borsboom et al. (2003, 2004), as well as Borsboom and Wijsen (2016), make a valuable contribution to this debate by clearly articulating criteria for scientific validity relating to the nature of the latent variable and the consideration of ethics. There are branches of psychological assessment that can fulfil these criteria including the behaviourist and biological approaches (Fernald, 2007; Pavlov, 1927; Skinner, 1938, 1957). However, the

field of educational assessment is not able to fulfil these criteria. One response to the challenge by Borsboom et al. (2003, 2004) is therefore that educational assessment forges its own identity independent of psychological assessment.

There are two broad imperatives to differentiate validity into two paradigms. The first imperative emerges from the growing interest in psychology and particularly biological psychology. Habermas (2003/2015), for example, in *The Future of Human Nature*, addresses developments subsequent to the separation of the elementary components of the genome, prenatal diagnosis, and in vitro fertilisation. Psychological approaches that address these types of physiological developments are more consistent with approaches of realism. Second, this study finds developments in biology as distinct from developments in knowledge-based economies and their relationship to educational outcomes (Drucker, 1968/1992; Hanushek, 2016; Hanushek & Ettema, 2017). While these two branches both address human behaviour, each take a distinct path to suggest that tensions will increase within the current paradigm that embraces both the biological and sociocultural paths. The following section continues to explore these pressures on validity paradigms through the concept of performativity.

7.3. Performativity and the performative attitude

This section addresses the concept of performativity, as elaborated by Lyotard (1979/1984), and juxtaposes performativity to the performative attitude described by Habermas (1981/1985) and Weir (1995). The concept of performativity relates to the realist stance through its relationship to behaviourism and empiricism. Skinner (1938), for example, introduces the notion of conditioning for both reinforcing and extinguishing certain behaviours. Education's concern for modifying behaviour and performance can be couched in similar terms, as do Engelmann et al. (1988) through the pedagogy of direct instruction that includes techniques of reinforcement and correction. The behaviourism of Skinner (1938) also draws on concepts of maze learning (Hull, 1932), and this study suggests that maze learning correlates to aspects of performance and performativity. Performativity is characterised here as instrumentally focused and not specifically engaged with the three worlds and substantive meaning in a way that contrasts to the performative attitude.

Performativity, four aspects

Lyotard (1979/1984) describes performativity in the field of knowledge production and knowledge reproduction through four broad characteristics. The first characteristic is that performativity pertains to activity that does not specifically address what is true, just or beautiful, or does not specifically address the pattern of three worlds explored in Chapter 4. Instead, performativity refers to activity that addresses what is efficient, useful and saleable. The second characteristic is that performativity involves activity where forms of communication are restricted in a way that privileges the language game of science and the language games of the powerful (Wittgenstein, 1953/2009). The third characteristic is that performativity pertains to systems that define what kind of performance is valued and how much it is valued. In doing so, performative systems create identities that reflect the kind of performance valued by a system. This characteristic is explored in this study through work of Austin (1962/1975) on speech theory and work of Butler (1990/2007) on identity construction. The fourth characteristic of performativity described by Lyotard (1979/1984) is that it refers to activity influenced and motivated by commercial interests.

The concept of performativity emerges from the work of Austin (1962/1975) on how words and language can be used to do things that create the social world. Lyotard (1979/1984, p. 88) uses the notion of the performative to describe systems that have input/output efficiency ratios. In the foreword to Lyotard's (1979/1984) book, Jameson (1984) associates performativity as scientific work not just tasked with replicating and modelling an aspect of reality, but to simply produce more work and to generate new and fresh scientific statements. Ball (2003) introduces the concept of performativity to the field of educational research, after which it is often used to describe effects of educational assessment as described in Chapter 2 (Hardy & Lewis, 2016; Keddie, 2016; Lambert et al., 2015; Solomon & Lewin, 2016). Performativity is here described through four aspects that apply to education, and which primarily draw on Austin (1962/1975), Lyotard (1979/1984) and Butler (1988, 1990/2007). Examples are also drawn from the field of educational assessment.

One aspect of performativity is its focus on instrumental reasoning rather than a focus on the objective, social and subjective worlds. A focus on instrumental efficiency is evident

in human capital theory (G. S. Becker, 1964/1993). Hanushek (2016), and Hanushek and Ettema (2017), explore efficiency ratios and productivity ratios using educational assessment data without referring to content, or without referring to notions of truth, justice and beauty. This study also associates these critiques with neoliberalism and monetarist economics (Friedman, 1953/2008, 1969/2008, 1962/2008). This is not to suggest that monetarist economics drives performativity, but to suggest that neoliberalism is a correlate of performativity through its concern for efficiency and productivity while eschewing consideration of truth, justice and beauty.

The instrumental aspect of performativity also affects how analytic and substantial arguments apply to education (Toulmin, 1958/2003). When educational performance is increasingly valued in relation to educational assessment data, such as those produced programs such as NAPLAN and the PISA, educational arguments become increasingly analytic. That is, educational arguments become grounded in static data sets and informed by data analysis of available data, to render silent arguments on what is true, right and beautiful. Performativity favours analytic arguments at the expense of substantial ones (Toulmin, 1958/2003).

A second aspect of performativity is that it restricts forms of expression. Lyotard (1979/1984), after Austin (1962/1975), associates performativity with systems that define what is considered performance and define what is considered good performance. When performance is considered in terms of data sets such as those of NAPLAN and the PISA, valued performance is restricted to the performances described through these databases. Lyotard (1979/1984) draws on the concept of the language game developed by Wittgenstein (1953/2009) to associate performativity with a “set of rules one must accept in order to play the speculative game” (p. 39). The language game of science extinguishes the language of normative prescription, with Lyotard (1979/1984) arguing that

science plays its own game; it is incapable of legitimating the other language games. The game of prescription, for example, escapes it. (p. 40)

Performative systems privilege the language game of science in defining and specifying what forms of social expression are allowed. By circumscribing how citizens can express themselves performative systems also define the nature of social bonds (Lyotard, 1979/1984; Wittgenstein, 1953/2009).

Lyotard (1979/1984) evokes the notion of terror in relation to language games and the nature of social bonds, where terror relates to

the efficiency gained by eliminating, or threatening to eliminate, a player from the language game one shares with him. (p. 63)

In these ways, performativity defines what kinds of expressions are expected, and threatens to exclude those who do not conform to accepted forms of expression. The notion of terror introduced by Lyotard (1979/1984) is adopted by Ball (2003) in education, through his seminal work titled *The Teacher's Soul and the Terrors of Performativity*.

Kress (2003), in developing the concept of enhanced semiotic resources, provides an insight consistent with Lyotard's (1979/1984) thesis, noting that the new affordances of technology

imply a radical social change, a redistribution of semiotic power, the power to make and disseminate meanings. (p. 17)

In educational assessment, semiotic resources are generally described in terms of item-types (Scalise & Gifford, 2006; Sireci & Zenisky, 2016). New item-types represent the semiotic resources for educational assessment, and these item-types have the potential to define performance and have the capacity to disenfranchise students. Students can be disenfranchised either by not being assessed through item-types with which they are familiar, or by being denied opportunity to become familiar with item-types in an assessment.

In the context of technology-based educational assessment, performativity can occur when the item-types and response types available within an assessment are restricted. Riconscente et al. (2016), for example, in the context of designing performance assessments, state that their ECD process emphasises the precise specification of what

work products will be, and narrow down the kinds of features that will be central or optional for grounding the targeted claims. (Riconscente et al., 2016, p. 56)

That is, ECD limits the response options available to students in an assessment consistent with the concept of performativity. However, this can also be a sensible design constraint in many cases given that item-types, and assessment in general, can be costly to develop and implement. This is to suggest that the motivations for performativity may not always be related to power, even though it may give rise to power relationships as suggested by Kress (2003) and Lyotard (1979/1984).

A third aspect of performativity is that performative systems can construct identity. As discussed in previous chapters, Sadler (2009, p. 159) observes that teacher judgements can be profound through their sense of permanence when reported through transcripts, certificates and qualifications. In this way, educational assessment can construct and reify student identity. The work of Butler (1988, 1990/2007) also suggests that educational assessment can socially construct identity. Issues around when and how to create an identity is further explored Chapter 8. Messick (1989), for example, considers traits as stable and enduring characteristics of a person, in contrast Moss (2003) argues for more transient interpretations in the classroom context. While this study finds these two positions are not necessarily contradictory, it is raised in this context to highlight the potential of systems inadvertently creating identity as illustrated in the study by Howell (2017).

The student responses observed by Howell (2017) in relation to identity can be somewhat explained by the work of Reay and Wiliam (1999), who explore the extent to which children's perceptions of tests contribute to their understandings of themselves. Reay and Wiliam (1999) observe differential dispositions among children on the extent that they consider "the testing process as a definitive statement about the sort of learner they are" (p. 343). Reay and Wiliam (1999) conclude that further investigation is required "to map out the extent to which both pupil and teacher identities and practices are being modified through new assessment processes" (p. 352). In addition to how students regard an educational assessment, the extent to which systems create identity can also relate to the illocutionary force with which feedback is provided. This is all to suggest that student reaction and responses to feedback from educational assessment are complex (Lipnevich et al., 2016).

A fourth aspect of performativity is its association with commercial ownership of systems. Lyotard (1979/1984) employs the term *mercantilization of knowledge*, with performativity affecting “the privilege the nation-states have enjoyed, and still enjoy, with respect to the production and distribution of learning” (p. 5). For Lyotard (1979/1984), the mercantilization of knowledge shifts the criteria in education from truth, justice and beauty to utility and saleability. The potential for private interests to influence was illustrated in Chapter 6 through the influence of proprietary university selection exams. There are ongoing critiques on the effect of commercial interest in education and educational assessment (Bulkley & Burch, 2011; Burch, 2009; Burch & Good, 2014; Hogan, 2016a; Hogan et al., 2015, 2016; Lingard, Sellar, Hogan, & Thompson, 2017).

Nevertheless, this study identifies that educational assessment systems, particularly large-scale assessments, generally display elements of the four aspects of performativity described here. Educational assessment generally focuses on a particular domain to thematically stress limited dimensions of knowledge. Further, every educational assessment inherently restricts how students can respond, inherently reports in a way that gives rise to both adaptive and maladaptive responses by students, and tends to have some commercial involvement (Hogan, 2016a; Hogan et al., 2015, 2016; Lingard et al., 2017). These effects relate to legitimacy and might be considered part of educational assessment validation (Kane, 2006).

Performative attitude

The performative attitude provides a contrasting perspective to performativity in the context of educational assessment. The distinction between the performative attitude and performativity is analogous to the distinction between the item paradigm and activity paradigm identified by Behrens and DiCerbo (2014) as described in Chapter 2. Through the item paradigm, questions are posed to which students provide responses that are interpreted based on correctness. In contrast, the activity paradigm requests actions from students that are interpreted through features using multidimensional analyses. Where the activity paradigm is somewhat associated with performativity and instrumental efficiency, the item paradigm is associated with the performative attitude that can explicitly engage the three worlds identified in Chapter 4.

Habermas (1981/1985) describes the performative attitude in terms of citizens creating their own identities through opportunities to provide “yes” and “no” responses to propositions put to them. The performative attitude is analogous to the *Socratic method* and *Socratic dialog* where the teacher asks questions of students and a method designed to allow students to create their own mind and identity (Delić & Bećirović, 2016; Ritchhart & Perkins, 2005). The role of the performative attitude in the creation of citizenship is consistent with Kress’ (2003) observation that “writing will remain the preferred mode of the political and cultural elites” (p. 1). The relationship between the performative attitude, writing, and society’s elites is illustrated, for example, through the use of the Socratic method in fields such as law (Szypszak, 2015). Writing, and the performative attitude, is likely to remain the preferred mode for elites because, unlike performativity, the performative attitude explicitly addresses intersubjectivity and expressions of identity through the three worlds.

For Habermas (1981/1985), the

performative attitude that ego and alter adopt when they act communicatively with one another is bound up with the presupposition that the other can take a "yes" or "no" position on the offer contained in one's own speech act. Ego cannot relinquish this scope for freedom even when he is, so to speak, obeying social roles; for the linguistic structure of a relation between responsible actors is built into the internalized pattern of behavior itself. (p. 59)

This suggests that the performative attitude is inherent to intersubjectivity. As soon as two actors interact communicatively, the performative attitude is required. Furthermore, the performative attitude provides students with an opportunity to take responsibility for a “me”, and to be recognised as an “I”. In this sense, the performative attitude is important for socialisation, individuation, and identity development. The notion of identity construction through communication is explored by Weir (1995),

The demand made of participants in linguistically mediated interaction, that they accept the freedom and responsibility of taking an affirmative or negative position in response to an other’s speech act, is what underlies

the development of moral identity as a critical relation to social norms, and of personal identity as a critical relation to oneself. (p. 273)

In this way, the performative attitude pertains not only to the cognitive dimension, but also to the moral and aesthetic dimensions. Enacting or performing responses to questions fosters the process of student individuation to construct identity.

The performative attitude in communication described by Habermas (1981/1985) and Weir (1995) gives rise to several corollaries in the context of educational assessment. One corollary is that a student's response to test questions forms their identity and is part of socialisation and individuation. This makes educational assessment, particularly in the form of the question–answer format, an integral part of the education process. As identified in Chapter 4, the purposes of educational assessment are therefore not solely exogenous to the classroom and student, and not solely for generating and extracting data for teachers and other stakeholders. Educational assessment is a form of communication in its own right, regardless of whether persistent data is created. Educational assessment, in presenting students with stimulus to which they need to provide a response, is an inherent part of teaching that fosters and constructs student identity and a student's sense of self.

This is not to suggest that the performative attitude does not result in an “outcome” or “output” of education. It is more to suggest that the performative attitude addresses the identity of the child as a project. As Oakeshott (1962/1991, p. 188) argues, the product of education is the activity invested in the child. The performative attitude directly addresses this activity by requiring students to respond to create their own identity. Educational assessment can engage a student in: the qualification purpose through how students might perform at work, the socialisation purpose through how students respond to societal norms, and the subjective purpose of the self through developing what students may enjoy and demand of others. The performative attitude addressed in educational assessment might also, but not necessarily, be captured in persistent data for broader societal coordination.

A second corollary is that educational assessment has the potential to distort the educational process through what is assessed. This occurs when the emphasis given by educational assessment programs do not appropriately reflect the normative cultural

context of the three worlds as described in Chapter 4. That is, when educational assessment does not provide adequate opportunities for students to engage and express themselves in these worlds, students may be denied opportunities to individuate along cognitive, moral or aesthetic dimensions. In terms of the exchanges that the school and the household have with the economy, a failure to individuate may lead to a failure to develop qualification attributes for work, a failure to socialise with societal norms, or a failure to create a sense of self. A failure of socialisation, or a poor sense of self and poor ego-strength, may arise, for example, when educational assessment places an excessive focus on the qualification purpose of education. The potential for educational assessment to distort education through the content it assesses therefore pertains to the concept of performativity.

A third corollary follows Habermas' (1992/1998) observation that the possibility to take "a yes or no position to a criticizable validity claim only [arises] if the other is willing to justify the claim" (p. 119). That is, where a student rejects a claim by a teacher, or provides an "incorrect" answer to a question, there needs to be an opportunity for validity to be intersubjectively redeemed between the teacher and student through discourse. In this sense, the question-answer format of educational assessment only signals the start of the interaction. The full validity of the interaction, particularly where a student answers "incorrectly", is redeemed through discourse addressing the objective, social and subjective worlds as appropriate through processes of justification and application (Habermas, 1994).

While Habermas (1981/1985) casts the performative attitude in "yes" and "no" positions, Weir (1995) casts these in terms of an affirmative or a negative position. In the context of complex issues, the performative attitude might therefore be expressed through extended constructed responses, particularly essays that require students to take a performative attitude. Essay writing has various genres such as descriptive, narrative, expository, argumentative, instructional, and transactional (OECD, 2010a, pp. 32-34). While this study does not seek to provide commentary on what essay types and prompts are most appropriate for engaging students in the performative attitude, Chapter 2 described the availability of automated essay scoring (Cohen & Wollack, 2006; Shermis et al., 2016). McCurry (2010) also describes how automated scoring may be more appropriate for some essays prompts than others, which suggests that automatic scoring

involving proprietary algorithms can curtail essay types and essay prompts. Therefore, while essays can provide opportunities to take a performative attitude, essays can drift into performativity when expression is restricted or defined through commercially developed algorithms. These tensions around the performative attitude and performativity might be addressed through educational assessment validation (Kane, 2006).

The performative attitude beyond the cognitive

This study observes the performative attitude in the cognitive dimension as the dominant focus of contemporary educational assessment. However, in a manner consistent with the broadening of progress narratives theorised by Habermas (1976/1979), there is an increasing interest in assessing the moral-practical and aesthetic-expressive dimensions. This is to suggest an increasing focus on student individuation and the performative attitude with respect to all three worlds.

The PISA provides a clear case example where the cognitive dimension is explored along with other dimensions. The PISA has included the affective dimensions in scientific literacy since 2006 (OECD, 2006, 2007, 2009a). In addition to cognitive aspects of scientific literacy, the PISA assesses

non-cognitive aspects: how students respond affectively. Attitudinal aspects of their response engage their interest, sustain their support, and motivate them to take action (Schibeci, 1984). Through such considerations we are led to define the overarching domain of *scientific literacy* for PISA 2006. (OECD, 2006, p. 22)

This provides a clear case contemporary example of educational assessment practice where, in Habermas' (1981/1985) terms, the objective, social and subjective worlds are thematically stressed within the assessment of a single domain. This study does not make an argument on how the affective domain maps onto each of the three worlds. Instead, the assessment of scientific literacy in the PISA is cited to provide clear case support that the social and subjective worlds are of interest in contemporary educational assessment design and use.

For the social and subjective worlds, issues of race and sexuality are also emerging as increasingly important as part of the performative attitude. Schilt and Bratter (2015), for

example, explore multiracial and transgender categories in the census in the United States. While simply having an identity recognised is a rudimentary form of assessment, Schilt and Bratter (2015) alert to issues involved in even these small steps. Further, an increased focus on the social and subjective worlds is also likely to increasingly affect how educational assessment conceptualises cognitive domains. This follows from Habermas' (1981/1985) claim that all three worlds are simultaneously engaged in communicative action, an issue addressed by Fensham (2016) in relation to gender and the assessment of science in Australia.

Broader interest in the social and subjective worlds of students is also demonstrated by the OECD with Kautz et al. (2014) describing the role of non-cognitive and personal traits in relation to work in the economy. This broader interest is also explored by Farrington et al. (2012), further addressed by this study in Chapter 8 in relation to the self.

This section has identified a tension between performativity and the performative attitude, however this study does not make a claim as to what approach to educational assessment is more suitable for particular contexts. Instead, this study suggests that the tension between performativity and the performative attitude is a matter related to legitimacy, and a matter for educational assessment validation (Kane, 2006). This tension also relates to progress narratives where the databases of performativity tend to focus on a fix view of the world and performative attitude more responsive to the emerging world of the student.

The performative attitude, and the increasing interest in the affective domain which addresses the social and subjective worlds, provides further clear case support against the realist stance in educational assessment. The realist stance is more consistent with notions of performativity and performance assessment and its relationship to behaviourism (Hull, 1932; Skinner, 1938), as these forms of assessment do not inherently engage with the social and subjective worlds. These themes will continue to be explored in the next section in the context of educational assessment for management.

7.4. Educational assessment for system management

In 2005 Geoff Masters (2005), then Chief Executive Officer of ACER, in an article for *The Australian* newspaper, defended outcome focused education in the face of criticism that it was leading to a decline in achievement. Masters (2005) argued that the teaching

of a syllabus was no longer enough, and that a focus on outcome measures promoted system improvement. That is, Masters (2005) favours the use of output measures for system management rather than a focus on inputs such as curriculum, syllabus and teacher training. This suggests that Masters' (2005) advocacy for output measures is consistent with the realist stance through its similarly diminished concern for educational content.

Masters' (2005) defence of outcome focused education raises the issue of how educational assessment data is used to coordinate and manage education systems. This study identifies two broadly distinct forms of system management: strategic management and administrative management. Strategic management focuses on outputs and numerical coordination, and administrative management focuses on inputs and on linguistic coordination. The following discussion locates Masters' (2005) call for the use of output measures with strategic management (associated here with scientific management) (Jelinek, 1979; Taylor, 1911/1998), which is juxtaposed to administrative forms of management (Fayol, 1949/2013; Weber, 1946/2009a).

Strategic management

Masters' (2005) call for the use of output measures is consistent with scientific approaches to management generally credited to Taylor (1911/1998). It is an approach also critiqued by Eacott (2017) who considers the work of Hattie (2009), with its focus on effect sizes identified through meta-analyses of research studies, as similarly oriented towards Taylor (1911/1998). The era introduced by Taylor (1911/1998) is characterised "by a search for efficiency and systemization of management thought" (Wren, 1994, p. 103). Where earlier approaches to management focused on authority through rules of bureaucracy (Weber, 1946/2009a), Hofstede (1994) describes that "Taylor was not really concerned with the issue of authority at all; his focus was on efficiency" (p. 8). Masters' (2005) argument that education system management similarly focus on output measures, and not inputs such as curriculum and syllabus, is consistent with these aspects of Taylor's (1911/1998) work.

The outcomes approach of Masters' (2005) is also consistent with approaches to strategic management and planning (Mintzberg, 1994). Jelinek (1979), in part, credits Taylor (1911/1998) for the emergence of strategic management. The rise of strategic management can be attributed to two management developments:

first is an attempt to rise above the concrete details of the task to think about what is done, rather than merely do it ...

The second trend visible in the development of management is a continuing attempt to transcend dependence upon the skills, memory, or capacity of any single individual. (Jelinek, 1979, p. 8)

These two trends are consistent with developments in education system management. Callahan (1962/2007), for example, associates the emergence of school administrators and school executives with Taylor's (1911/1998) scientific management. Callahan (1962/2007, pp. 215-219) observes that the profession of school administrator became established around 1918, a development influenced by businesses in general adopting Taylor's (1911/1998) approach. These developments gave rise to a bifurcation of education into two broad professions, one of teacher, the other of administrator. Callahan (1962/2007, pp. 153-158) also describes records and reports as mediating between these two professions, in a way that "large corporations regarded their records of the manufacturing process" (p. 153).

The outcome and data driven approach of Masters (2005) is also consistent with contemporary critiques over efficiency and productivity in educational assessment. Hanushek and Ettema (2017), for example, use outcome measures in the form of educational assessment data to explore the efficiency of education systems. This approach is consistent with observations by Callahan (1962/2007) that developments in scientific management led to a concern for "cutting costs and of demonstrating efficiency" (p. 158). These developments are all consistent with Jelinek (1979), who describes strategic management as linking accounting and other information through unified control systems and into a unified theory of management. Furthermore, in linking measured student outcomes to accounting, the strategic management focus is consistent with human capital theory (G. S. Becker, 1964/1993). This abstraction also demarcates strategic management from operations management in work units consistent with the outcome focused approach advocated by Masters (2005).

Strategic management in education

The data-driven and outcome focused approach advocated by Masters (2005), and others such as Hattie (2009), is consistent with strategic forms of system management described by Jelinek (1979) and illustrated in Figure 12 below.

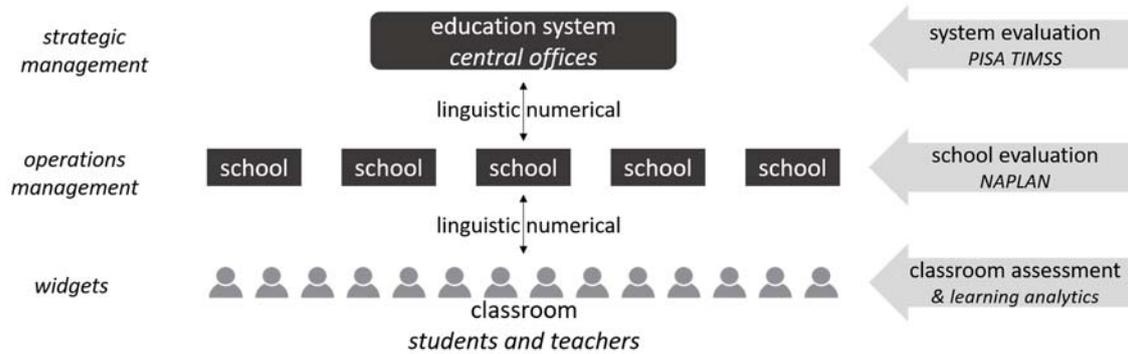


Figure 12 – Strategic management in school systems

The key feature of Figure 12 is that it shows the division between teachers and administrators that Callahan (1962/2007) associates with the rise of scientific management. The distinction between teacher and administrator is reflected in the respective distinction between operational management and strategic management described by Jelinek (1979) and illustrated on the left of Figure 12. The middle of Figure 12 reflects the three levels of data use, as described in Chapter 4 which draws on the OECD’s (2013c, p. 60) report *Synergies for Better Learning*. The three arrows on the right of Figure 12 show some of the data available for management in educational systems at the various levels and is illustrated with reference to Australia’s NAPLAN. Figure 12 relates to this study’s broader concerns on how the symbolic media of educational assessment is used across the levels of the system. In particular, Figure 12 exposes the question over the extent to which both the linguistic and numerical components are communicated across the system.

Figure 12 illustrates that when the outcome focused approach advocated by Masters (2005, 2013) is applied to system management, the potential for a decoupling of the linguistic and numerical component of educational assessment emerges. When education system performance is evaluated through programs such as the PISA and the TIMSS, some interpretation and translation may be required as these programs might not fully

reflect local curricula. Figure 12 illustrates how schools can use these data from system evaluations, school evaluations, classroom-based assessment, and learning analytics to coordinate classroom activity. The question that emerges from Figure 12 is how well the linguistic component of educational assessment gets communicated and translated across each level along with its numerical component.

The question of how linguistic expectations arising from educational assessment are communicated across the system is heightened given that the evaluation regimes illustrated in Figure 12 refer to different educational content. Wu (2009) finds that the TIMSS and the PISA assess different content (Klieme, 2016). In the Australian context, Gorur and Wu (2015) find that the content in the PISA is different to that of, for example, the Australian curriculum. Furthermore, NAPLAN does not claim to assess curriculum, with its tests designed to only “broadly reflect aspects of literacy and numeracy within the curriculum in all jurisdictions” (ACARA, 2015b, p. iv). This is to raise the question on how the linguistic component of educational assessment is translated and communicated across systems.

The literature suggests that the linguistic component is not translated across education systems, and that only the numerical component is communicated. This development is consistent with Masters’ (2005) call for the use of output measures to coordinate education systems. Biesta (2015b) frames this as “the seductive nature of numbers” (p. 348) in how education systems work and perform. Grek (2009) describes this in terms of “governing by numbers”. Gillis et al. (2016) express concern over how educational discourse is influenced by think tanks who “interpret, manipulate and disseminate the results of data arising from large scale assessment survey programmes such as PISA” (p. 132). Ball (2015) refers to this as the “tyranny of numbers” in education governance. Lingard (2011) refers to this as “policy as numbers”. This is to suggest that Masters’ (2005) advocacy of output measures for system management is a dominant paradigm in contemporary education. It also suggests that it is an approach that is the subject of extensive negative critique, which is in effect a critique of numerical coordination across education systems.

A case emerges that the outcomes focus advocated by Masters (2005, 2013), and its associated strategic approaches to management (Jelinek, 1979; Taylor, 1911/1998), have

led to a decoupling of the linguistic and numerical components of educational assessment. Masters (2005) explicitly rejects curriculum and syllabus as being the primary drivers for coordinating education activity, and instead advocates for output measures to coordinate systems. In the context of Australia this reflects a shift from coordination through curriculum (Barcan, 1996, 2003), to coordination through the numerical component of educational assessment. In defending the outcome focus in system management, Masters (2005, 2013) also identifies it as the dominant approach in Australian education. This is to suggest that it may also provide an explanation for Australia's struggles on the PISA scales (Thomson et al., 2013; Thomson et al., 2017).

The numerical and outcome focus advocated by Masters (2005, 2013) is consistent with the market driven approaches of neoliberalism in managing complexity in society (G. S. Becker, 1964/1993; Friedman, 1969/2008, 1962/2008). It is an approach that is supported and enabled by a realist stance towards the latent variable of educational assessment that considers traits to exist independently of what is said about them (Borsboom et al., 2003, p. 209). This approach also turns away from progress narratives through eschewing discourse over content. In the absence of any translation of the linguistic components between measurement regimes and local curriculum, the realist stance provides the only way of interpreting numerical data.

The outcomes focus advocated by Masters (2005, 2013) promotes and facilitates a realist stance toward the latent variable. It also facilitates market driven approaches that render what Friedman (1962/2008) describes as an "approved educational services" (p. 89), unproblematic and unable to be challenged. It renders any challenge to curriculum on the basis of gender silent (Fensham, 2016). In this way, it potentially reinforces white male privilege (McIntosh, 1986), and renders any challenge on the basis of feminism as challenges to authority as described by Caldwell and Hayward (1998, p. 12). In this way, the outcome focused numerical neoliberal approach to system management counters progress on social matters.

Administrative management as an alternative

This study identifies administrative management as a form of system management that contrasts to the outcomes focus advocated by Masters (2005, 2013), and its associated scientific and strategic management approaches (Jelinek, 1979; Taylor, 1911/1998).

Administrative management is associated with Fayol (1949/2013) and is a form of management generally concerned with the organisation of inputs into organised work. Hofstede (1994), for example, associates the approach of Fayol (1949/2013) with the approach to bureaucracy described by Weber (1946/2009b). Weber (1946/2009b, p. 196) describes bureaucracy as having features such as fixed jurisdictional areas governed by rules, which gives rise to authority to give commands, where authority is also governed by rules. These features of bureaucracy and administrative management provide an overall coherence and unity to coordination and management.

The approach to management advocated by Fayol (1949/2013) is also concerned with a unity of command. Hofstede (1994, p. 9) describes Fayol (1949/2013) as generally impressed by the technical output measures approach of Taylor (1911/1998), but shocked by its disregard for the principle of unity of command. It is through the focus of unity of purpose that administrative management differs from the decoupled output focus advocated by Masters (2005, 2013).

Through its support for procedural rules and unity of purpose (Fayol, 1949/2013; Weber, 1946/2009a), administrative management provides better support for the principles of discourse ethics (Habermas, 1983/1996) and deliberative democracy (Habermas, 1992/1998). Administrative forms of system management are better at supporting an integrative system with the capacity to institutionalise postconventional reasoning (Kohlberg, 1971) and principles and procedures for universal inclusion (Habermas, 1983/1996). For these reasons, administrative management and bureaucracy is also eschewed by advocates of neoliberalism (Friedman, 1962/2002; Hayek, 1944/2007; Popper, 1945/2002; von Mises, 1944/2007).

Administrative management, when applied at the system level, promotes unity across society and is distinct from the strategic management of neoliberalism which Habermas (2001) considers as scattering the lifeworld into distinct groups and monads. A legitimisation tension therefore emerges as to which management approach most effectively and efficiently meets the needs of society and its citizens.

A postmodern rejoinder

The postmodern writer Baudrillard (1981/1994) provides an interesting counterpoint in relation to the symbolic media used for communicating between strategic and operational

levels illustrated in Figure 13. It is an observation that also pertains to the idea of Parsons (1963a) that media involves the symbolic to be accepted in lieu of the real. Baudrillard (1981/1994), in his book *Simulacra and Simulation*, describes how the symbolic becomes disengaged from reality, to turn it into a simulacrum that bears no relationship whatsoever to reality. This study finds some supporting arguments for this observation in educational assessment. That is, educational assessment can be used in discourse and decision-making while bearing no relationship to any reality, and that there is an inherent tendency for slippage between the reality of educational activity and educational assessment as symbolic media that represents it. This slippage is consistent with the position of Baudrillard (1981/1994).

Powell (2012) observes that Baudrillard (1981/1994) is “often seen as the ‘high priest’ of postmodernism” (p. 8), and associates Baudrillard (1981/1994) with a distrust of progress narratives. While pragmatism similarly identifies an inherent tension around how well educational assessment reflects the reality of educational activity, it also considers these tensions as leading to legitimation crises. Following Habermas (1976/1979, 1975/2005), pragmatism theorises legitimation crises as manifesting in a way that causes societies to address them through ethical and moral development. Nevertheless, Baudrillard (1981/1994) provides an interesting counterpoint.

Luhmann (1984/1995) provides another counterpoint. Habermas (1981/1985) largely rejects the systems theory of Luhmann (1984/1995) that considers systems as autopoietic and self-propelling. Where Habermas (1981/1992) considers the action orientation of actors in the lifeworld, Luhmann’s (1984/1995) perspective totally focuses on the system. For Habermas (1981/1985)

Systems theory replaces “subject” with “system,” “object” with “environment,” and conceptualizes the subject's ability to know and to deal with objects as systemic achievements that consist in registering and reducing the complexity of the environment. (p. 396)

Luhmann (1984/1995) also provides a contrasting approach towards symbolic media, as representing non-random variation in social relations (Chernilo, 2002, p. 433). In the systems theory of Luhmann (1992), understanding

is never the mere duplication of the utterance in another consciousness but a condition of connection with further communication in the communication system, that is, a condition of the autopoiesis of social systems. (p. 253)

In this way, Luhmann (1992) is not so much concerned with redeeming the validity of social interactions. Instead, even if communication is misunderstood, it is either accepted or rejected by the receiver to become part of the system's autopoiesis and "the recursive operations of self-referential systems" (Luhmann, 1984/1995, p. xx). For Luhmann (1992), misunderstanding becomes just another element that produces and reproduces the system. Luhmann's (1984/1995) systems theory therefore contrasts to the approaches of both Habermas (1981/1992) and Baudrillard (1981/1994), in that the autopoiesis of systems has no concern for symbolic media to correspond to a reality, it is primarily concerned with maintaining the system.

Luhmann's (1984/1995) perspective also suggests that the approach of Habermas (1981/1992) may be somewhat aspirational in describing how system generated symbolic media *ought* to be generated rather than how systems actually generate symbolic media for society. This aspirational aspect is acknowledged by Habermas (1983/1996) through his notions of ideal speech acts, and ideal role taking. Luhmann (1992) does not share similar idealistic concerns for truth and rightness.

Closing remarks

This chapter opened the case that educational assessment should be considered distinctly from psychological assessment. Where psychological assessment often addresses scientific and ethical matters simultaneously, it is also a field that seeks to generalise beyond cultural horizons, particularly through biological and behaviour approaches (Fernald, 2007). In contrast, valid use of educational assessment is oriented towards a sociocultural context and is responsive to dynamics around technological development and social progress. Psychological assessment does not always share these concerns.

In distinguishing educational assessment from psychological assessment, the connection between the lifeworld of the classroom and how that is reflected in the design and use of educational assessment is foregrounded. This particularly pertains to how progress narratives are addressed. The postmodern perspective of Baudrillard (1981/1994) and the

systems theory of Luhmann (1992) are particularly useful to illustrate this point. Baudrillard (1981/1994) describes media as developing in a way that it no longer reflects any reality, and Luhmann (1992) describes system media as not being concerned with matters of truth and rightness. These attitudes contrast to the pragmatic approach adopted by this study, which seeks to ensure that educational assessment remains relevant to the students as they emerge as citizens in society.

Chapter 8. Towards a reconstruction of educational assessment validity

This chapter begins to reconstruct approaches to educational assessment. The first section addresses normative validity (Biesta, 2009) and describes educational assessment as an activity that communicates societal values and expectations (Mislevy, Steinberg, & Almond, 2003). A number of implications and aspects of normative validity are then discussed, including processes of justification and application (Habermas, 1994; Rehg, 2011), permanence and transience (Messick, 1989; Moss, 2003), and semiotic resources (Kress, 2003, 2010). An issue with multiple comparisons in large-scale educational assessment is also explored to provide a scientific justification that further information is required for valid actual interpretation in local contexts.

The second section explores how the PISA addresses normative validity. The section speculates on how the PISA attained its legitimacy as one of the most influential international assessment programs (Biesta, 2015b; Bonal & Tarabini, 2013; Breakspear, 2012; Grek, 2009). It also speculates on why the legitimacy of the PISA might be eroding ("OECD and PISA tests are damaging education worldwide – academics," 2014).

The third and final section works towards answering the study question on how different attitudes towards progress narratives affect educational assessment design and use. It summarises some of the legitimisation tensions identified through this study, cast as a tension between pragmatic approaches that engage with progress narratives and those, including neoliberalism, that eschew them. These are enumerated with the view that each might be addressed, as appropriate, through the better argument in the design and use of educational assessment to ensure that the needs of society and its citizens are best met.

8.1. Normative validity

The sociocultural orientation of education brings to the fore the notion of normative validity. Braaten (1991, p. 31) describes normative validity claims as raised indirectly in the course of daily interaction. Whenever normative validity claims are challenged in daily interaction, practical discourse arises that simultaneously address claims to truth, normative rightness, sincerity and truthfulness. Biesta (2009) argues that normative validity is being replaced by technical validity in educational assessment and questions “whether we are indeed measuring what we value” (p. 35). Two further questions emerge

from this claim by Biesta (2009). One relates to who are the “we”. The second relates to how does this “we” determine what “we value”. It is to these questions that Habermas’ (1983/1996) principle of universalisation (U), and the principle of discourse ethics (D), provides a useful framing.

Biesta (2009) associates technical validity with the “rise of a culture of performativity in education” (p. 35). As argued in Chapter 7, Lyotard (1979/1984, p. 18) associates performativity with an emphasis on denotative statements that exclude human interests such as justice, happiness, beauty and wisdom. Performativity does this by focusing on performance as defined by systems and in a way that prevents validity claims from being redeemed through communicative action (Habermas, 1981/1985; Lyotard, 1979/1984, p. 88). Where performativity focuses on features and efficiency of performance (Behrens & DiCerbo, 2014; Messick, 1994; Mislavy, Almond, & Lukas, 2003), communicative action focuses on linguistically redeeming normative validity claims and meaning (Habermas, 1992/1998; Weir, 1995). This section describes some issues associated with normative validity.

Normative validity

Normative validity is generally expressed through curriculum and syllabus. As described in Chapter 2, this study finds that complexity arising from negotiating and harmonising issues of curriculum from the 1960s through to the 1980s in Australia led to changes in how education systems became coordinated. This study observes that Australia’s education activity was linguistically coordinated through curriculum emerging from the 1960s to the late 1980s, and numerically coordinated through educational assessment from the early 1990s onward (Barcan, 1996, 2003; Masters, 2005). While there is an argument that this transfer was ideologically driven by the rise of neoliberalism (Marginson, 1993, 1997a), the accounts of Barcan (1996, 2003) suggest that curriculum reform, particularly through the 1980s, was becoming increasingly complex and contested. Barcan (1996, 2003) describes an ongoing tension between school-based curriculum responsive to local contexts, and centralised curriculum responsive to demands of the national economy as expressed through human capital theory (G. S. Becker, 1964/1993; Marginson, 1989). Further, the account of McCurry (2017) on the history of test development in Australia, is consistent with increasing pressure on system

integration across Australian jurisdictions and school sectors leading to cross-curricular assessments linked through equating (Dorans et al., 2011; Holland & Dorans, 2006). In this context, Biesta's (2009) observation about a loss of normative validity in educational assessment is justified.

A loss of normative validity can be associated with symbolic media increasingly losing its connection to its linguistic meaning and transforming into abstract steering media. Connection can be lost through processes of equating and linking (Dorans et al., 2011; Holland & Dorans, 2006) associated with item response theory as described in Chapter 2 (Hambleton et al., 1991; Wu et al., 2017). Equating and linking can extend the validity of the numerical component of educational assessment to horizons beyond which its linguistic component is normatively valid. Chapter 7 illustrated this with respect to outcome focused system management (Masters, 2005), where the numerical component of programs such as the TIMSS, the PISA and NAPLAN, which do not directly reflect local curriculum, can be used without reference to the linguistic component to coordinate and manage activities of school systems. Mislevy, Steinberg, and Almond (2003) describe educational assessment as communicating societal values, expectations and norms. The outcome focused approach towards system management advocated by Masters (2005) illustrates how quantitative media, and the numerical component of educational assessment, is increasingly used to set norms and expectations in a way where these expectations become decoupled from linguistic meaning.

This study identifies an enduring dialectic between educational assessment and curriculum, and that an increasing focus on the numerical component of educational assessment draws away from substantive meaning expressed through curriculum. McCurry's (2017) account of test development in Australia provides a sense of the dialect that educational assessment has with teaching and curriculum, in suggesting that it

is probable that the teaching and examining in English during the 1970s was affected by the test papers in written expression and the humanities, which were widely studied by Australian educators. (p. 13)

Educational assessment, particularly since the early 1990s, has increasingly become the vehicle for setting norms and expectations in Australia rather than curriculum (Caldwell & Hayward, 1998; Masters, 2005).

Nevertheless, there is probably also an ideological element to how expectations are set through educational assessment. For example, the dialogues in Plato's (1963) *Republic* could be considered archetypal dialogues for contemporary state-sponsored curriculum committees. Popper (1945/2002), an early associate of the neoliberal movement (Stedman Jones, 2012), takes an explicitly critical position against the collectivist tradition of Plato. However, as argued in Chapter 6, this study considers that this leaves what is considered an "approved educational services", "approved institutions" and "approved school", terms employed by Friedman (1962/2002, p. 89), as either undefined, arbitrary, or reified to a particular, and probably dated, set of norms. A focus on the numerical component of educational assessment relieves the burden on state-sponsored curriculum committees to engage in linguistic meaning-making, but in doing so adversely affects normative validity and the linguistic connection that educational assessment has to the sociocultural sphere of society.

Another issue for normative validity relates to a loss of social and cultural precision that arises when an educational assessment program broadens its horizons. This was one motivation, for example, for the stimulus-based cross-curricular testing described by McCurry (2017). As described in Chapter 6, the pressures on system integration resulting from increasing social integration as well as the impulse of citizens to broaden lifeworld perspectives, place pressures on normative validity. The literature review of Chapter 2, observed that the PISA "drops" items that are too culturally specific. Similarly, national assessments such as Australia's NAPLAN only broadly reflect the curriculum of the educational systems it covers (ACARA, 2015b, p. iv). There is therefore slippage between educational assessment and curriculum, and hence between educational assessment and societal norms, as the horizons of an educational assessment program broaden.

Normative validity emerges when the linguistic content of an educational assessment is developed to address the normative local context through deliberative means (Braaten, 1991; Habermas, 1992/1998). Habermas (1981/1985) describes normative validity as "an interest common to all those affected and thus to deserve general recognition" (p. 19). This is to suggest that normative validity in educational assessment is concerned with addressing content that is relevant and important for all students undertaking an assessment. Further, the greater the focus on the classroom context the greater the normative validity. In the context of the classroom, normative validity might be achieved

through educational assessment, as described by Cowie and Bell (1999), that is sensitive to the needs of individual students, their social context, and the subject matter being assessed.

When the local context is a jurisdiction such as the state of Victoria, normative validity might be achieved through the panel members that set, review and mark exams being drawn from a pool of teachers experienced in teaching the subject to be examined. This is the case for the VCE for example (VCAA, n.d-a). For national and international assessments such as NAPLAN and the PISA, normative validity emerges on how well the assessment instruments reflect: curriculum, work in the economy, and social norms in respective jurisdictions. Normative validity in relation to the PISA is addressed in the next section, with institutional processes related to normative validity including those related justification and application addressed first.

The integrative system and normative validity

The process of justification and application is not new to educational assessment. Discourses of justification and application provide for reciprocity that allow for norms centrally established to respond to how these norms affect citizens in application to thereby enhance normative validity. As described in the literature review in Chapter 2, Baird et al. (2000), in the context of standard setting, highlight the need for bodies to justify and defend their decisions given their value-laden nature. The VCE in Victoria also operates on a principle of reciprocity involving discourses of justification and application, with each year's exams made publicly available along with reports from examiners (VCAA, n.d-b). These features, along with the participation of teachers in exam setting and marking panels, illustrate a rudimentary architecture for cyclic justification and application. The PISA similarly justifies its results, through releasing technical reports (Adams & Wu, 2002; OECD, 2005b, 2009a, 2012, 2014, 2017e) and assessment frameworks (OECD, 2000, 2003, 2006, 2010a, 2013a, 2017b). These examples illustrate of discourses of justification and application apply to the design and implementation of educational assessment.

The processes of justification and application (Habermas, 1994) contrast to the process of stakeholder management (Donaldson & Preston, 1995; Eskerod & Huemann, 2013). Arnstein (1969), for example, develops a spectrum of citizen participation; from

manipulation, therapy and informing; to partnership, delegated power and citizen control (Hart, 1992). While the process of justification and application does not directly map onto Arnstein's (1969) framework, the framework does highlight how citizen participation can be communicative through justification and application, or strategic and goal-oriented through manipulation and informing.

In the pragmatic tradition, the process of justification and application (Habermas, 1994) are functions of the integrative system (Parsons & Smelser, 1956/2005) of society, where it is addressed function through deliberative processes of government (Habermas, 1992/1998; Weber, 1946/2009a). In contrast, the neoliberal tradition of markets (Friedman, 1962/2002) might seek address processes of justification through processes of marketing (Kotler, 1972; Kotler & Levy, 1969) and stakeholder management (Donaldson & Preston, 1995; Eskerod & Huemann, 2013). Following Habermas (1983/1996), complex societies advancing in both technological and social dimensions require formal procedures for discourse and formal procedures of justification in developing both curriculum and the content of educational assessment. These are inputs into education that, as described in Chapter 7, might be better addressed through administrative forms of management that have regard for unity of purpose through legislated rules of bureaucracy (Fayol, 1949/2013; Weber, 1946/2009a).

Nevertheless, as addressed in Chapter 6, deliberative processes can only improve on market-based processes when participants in decision-making have adequate postconventional reasoning capabilities. Engagement at the lower conventional level is likely to result in high levels of contestation, similar to those described by Barcan (1996, 2003) as occurring in Australia during the 1980s. That is, the integrative system as part of government can only improve upon the strategic coordination of markets if government systems can reason in a postconventional manner so that the perspective of all citizens in society are duly considered in deliberations with decisions publicly justified.

Permanence and transience

If, when and how to construct identity through the performative function of educational assessment is an issue that emerges from the literature. A tension becomes evident when an observation by Moss (2003) is juxtaposed to one by Messick (1989). Moss (2003) observes that as a classroom teacher,

I have no need to draw and warrant fixed interpretations of students' capabilities; rather it is my job to help them make those interpretations obsolete. (p. 16)

In contrast, Messick (1989) describes responses to an educational assessment as samples of a person for whom traits are considered

a relatively stable characteristic of a person – an attribute, enduring process, or disposition – which is consistently manifested to some degree when relevant, despite considerable variation in the range of settings and circumstances. (p. 15)

These contrasting positions of Moss (2003) and Messick (1989) on the nature of the latent variable of assessment are not irreconcilable, but it does point to a broader issue explored here under the rubrics permanence and transience. This issue is important from the perspective of when and how to report test results which can construct identity.

The issue of permanence versus transience and normative validity pertains to the temporal moments in a student's education where it is appropriate to reify a student's identity. Following the three purposes of education described earlier in this chapter, this pertains to identity along the cognitive, moral or aesthetic dimensions, and where this identity might be constructed and reified through a teacher's utterance, grade or transcript. The notion of permanence and transience foregrounds the question as to which point in schooling a student's development might be translated into a socialised identity for use in a life-project as citizen. That is, at what points in schooling does the reporting of educational assessment data translate from communicative action that describes performance, to strategic action that creates identity.

As mentioned briefly in earlier chapters, there are two broad overarching approaches towards the temporal stability of the latent variable of educational assessment. One perspective is provided by Messick (1989, p. 15) who considers traits as a relatively stable characteristic of a person. This position contrasts to that of Moss (2003, p. 16) who considers educational assessment as not so much about fixed interpretations, but about making any interpretation redundant through further teaching and learning. The contrasting positions can be attributed to Moss (2003) taking a particular classroom focus

in her interpretation. Nevertheless, these positions illustrate a tension between permanence and transience in relation to educational assessment.

The notion of permanence has particular implications for the social and subjective world of students. Habermas (1981/1985), for example, considers expressive sentences as serving to truthfully manifest subjective experiences, he also considers that the sincerity of these expressions

cannot be grounded but only shown; insincerity can be revealed by the lack of consistency between an utterance and the past or future actions internally connected with it. (p. 41)

This creates a tension in the development of cognitive, moral and aesthetic identities among students. That is, where the process of education seeks to develop students along cognitive, moral and aesthetic dimensions, demands for sincerity require students to respond with a sense of consistency. For the cognitive-instrumental dimension, inconsistency is sometimes associated with *cheating* and as a dishonest or unfair act (Murdock, Stephens, & Grotewiel, 2016, p. 186). That is, a student performing on an educational assessment in a way that is inconsistent with past performance may lead to suspicions of inauthentic behaviour. Similar inconsistencies might be observed in trajectories of sexual identity of adolescents that might lead to claims of inauthentic behaviour (Savin-Williams & Diamond, 2000). Butler (1990/2007), for example, observes that

the “coherence” and “continuity” of “the person” are not logical or analytic features of personhood, but, rather, socially instituted and maintained norms of intelligibility. Inasmuch as “identity” is assured through the stabilizing concepts of sex, gender, and sexuality, the very notion of “the person” is called into question by the cultural emergence of those “incoherent” or “discontinuous” gendered beings who appear to be persons but who fail to conform to the gendered norms of cultural intelligibility by which persons are defined. (p. 23)

In the context of this study, what Butler (1990/2007) is suggesting is that there is a dialectical tension between social norms experienced by a student and their emerging

identity. This is to suggest a tension around truthfulness and sincerity as a student develops and individuates along cognitive, moral and aesthetic dimensions.

The work of Butler (1990/2007) alerts to the vagaries of identity construction in dimensions such as gender and sexuality. It reinforces the notion that the process of asking and responding to educational questions has a role in constructing identity and the self. Furthermore, these processes are dynamic and reflexive, where a student's moral and aesthetic identity evolves and develops with education (Kohlberg, 1971; Kohlberg & Hersh, 1977). Moral development therefore inherently involves instances of insincerity, as "a full grasp of moral conceptions must await maturity; the child's understanding is always primitive" (Rawls, 1971/1999, p. 403). This is to suggest that a student is likely to manifest a range of inconsistent moral, aesthetic and perhaps cognitive behaviours throughout schooling. It is through the process of teaching that inauthentic expressions – along cognitive, social and subject dimensions – are rendered obsolete as suggested by Moss (2003). This has implications for the reporting of educational assessment findings and results. It illustrates the need to identify when and how to report on development so that a student's cognitive, moral, and aesthetic identities are not prematurely reified through the performative processes of reporting.

The literature review of Chapter 2 described Tierney and Koch (2016, pp. 269-270) identifying a desire for privacy among students to facilitate development and learning. This study suggest that privacy also fosters healthier identity construction as privacy allows the fully formed identity of a student to emerge as a citizen without being hindered by incidental insincere expressions made during the years of schooling. How and when an educational assessment should affect the construction of identity is therefore a legitimacy concern that could be addressed through validation (Kane, 2006). Further, as identity, particularly sexual identity as described in Chapter 6, pertain to the tensions between the private world of the family and autonomous rights of the students, these questions also implicate the integrative system of society (Parsons & Smelser, 1956/2005).

Normative validity and semiotic resources

The enhanced semiotic resources afforded by technology also pose issues for normative validity (Kress, 2003, 2010). Habermas (1976/1979, pp. 9-12) considers understanding

as being based on a symbolically established intersubjective relationship with others. Habermas (1976/1979, 1981/1985) emphasises the importance of a “symbolically prestructured reality” (Habermas, 1981/1985, p. 107), and a “symbolic universe” (Habermas, 1976/1979, p. 88) to provide a basis for common understanding and communicative action. Habermas (1976/1979) considers that

universal claims (to the comprehensibility of the symbolic expression, the truth of the propositional content, the truthfulness of the intentional expression, and the rightness of the speech act with respect to existing norms and values) are set in the general structures of possible communication. (p. 97)

However, Habermas (1976/1979) frames communication in terms of speech acts and speech act theory (Austin, 1962/1975; Chomsky, 1969, 1957/2015; Wunderlich, 1979), and is largely silent on the notion of enhanced semiotic resources described by Kress (2003, 2010).

New validity and legitimacy issues arise when Habermas’ (1976/1979, 1992/1998) ideas of a symbolic universe and identity formation through the performative attitude is considered in terms of Kress’ (2003) notion of enhanced semiotic resources. Kress (2003, p. 1) considers technology as changing the nature of communication; from a world told, to a world shown, a change that involves major shifts in power relations beyond the field of communication. An issue that emerges relates to the level of opportunity students have to engage with contemporary semiotic resources normatively appropriate for an assessed domain, with some domains requiring specialised equipment (Navarrete, Santos, Hernández-Leo, & Blat, 2011; O’Loughlin, Ní Chróinín, & O’Grady, 2013; Penney, Jones, Newhouse, & Cambell, 2012). That is, it leads to a validity concern as to whether an educational assessment presents the assessed domain in an appropriate and authentic technological and semiotic form. In the field of educational assessment, semiotic resources are often considered in terms of item-types and item-formats (Scalise & Gifford, 2006; Sireci & Zenisky, 2016). As new semiotic resources emerge in educational assessment their use needs to be considered in discourses around validity and legitimacy (Kane, 2006). How new and emerging semiotic resources emerging through progress narratives are addressed in educational assessment pertains to validity and legitimacy.

Multiple comparisons and the normative context

Normative validity has so far been addressed through educational assessment's linguistic component, and how it promotes social and cultural reproduction. Reasoning from empirical science related to significance tests presents further argument that the numerical component of educational assessment is not capable of communicating scientific truth claims without further reference to linguistic and normative content. This is to further reinforce the need that actual interpretations in local contexts requires additional evidence drawn from the local context to redeem normative validity (Moss, 2016).

A central issue for comparative educational studies such as the PISA is that its findings and interpretations involve comparisons that have significance tests. How these comparisons are made and considered, and the nature of the significance tests, affects the validity of the comparisons. The PISA (OECD, 2016b, pp. 69-70), for example, reports country mean values on a scale with a 95% confidence interval. The table of country means is constructed in a way that each country's performance can be compared with other countries. Where confidence intervals do not overlap, various rankings can be ascribed. The key issue is that these tables provide for multiple comparisons between countries, without changing the confidence interval. Where the probability of a spurious conclusion increases whenever an additional comparison is made.

An issue with multiple comparisons arises with the PISA because while the

use of significance testing is an appropriate way to test a theory, ... as the number of comparisons increases so does the chance of false significance.
(Davies & Goldsmith, 1988, p. 379)

That is, the reporting of the PISA data does not consider, through more exacting significance criteria, the use of multiple comparisons in the interpretation of its data. Given the extensive number of reports and tables published for each cycle of the PISA, along with the option for secondary analyses, the number of possible comparisons is vast and beyond practical quantification (OECD, 2009b, 2016b, 2016c, 2017c, 2017d). It therefore becomes impractical to make the appropriate corrections to the significance tests, often referred to as a *Bonferroni correction* or *Bonferroni adjustment* (Collis & Rosenblood, 1985; Davies & Goldsmith, 1988; Dunn, 1959, 1961; Stanovich, 1988). The

following conclusion of Collis and Rosenblood (1985) therefore still applies to the PISA, and to much large-scale educational research, that the

problem of inflated significance in the analysis of a correlation matrix needs review. Neither the problem nor the approach recommended is new, but surprisingly little acknowledgment of either appears in the literature. Researchers using canned computer programs to analyse an intercorrelation matrix should make sure that a protection technique such as the Bonferroni procedure is used. The use of such techniques will reduce the likelihood of spurious effects being reported. (p. 55)

The PISA (OECD, 2016b, pp. 69-70) reports do not advise or promote adjustments to its significance tests. Instead, the manner of reporting promotes that each country be compared to all others using a 95% confidence interval without adjustment. This increasing the probability of spurious conclusions.

The approach in educational assessment contrasts to the approach taken by the pharmaceutical industry, for example, where for

confirmatory clinical trials, it is required to prespecify the outcome(s) to be evaluated, the hypotheses to be tested, and the analyses plan to be conducted. It has been a standard that a single primary endpoint is considered in a clinical trial. (Shi, Pavey, & Carter, 2012, p. 300)

Here, Shi et al. (2012) are drawing on the guidelines from the U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research, and Center for Biologics Evaluation and Research (1998). These guidelines state that a written protocol should be developed before a trial begins, and in order to

avoid multiplicity concerns arising from post hoc definitions, it is critical to specify in the protocol the precise definition of the primary variable as it will be used in the statistical analysis. (p. 7)

Shi et al. (2012) go on to propose adaptations to the Bonferroni correction using simulation. This rigour around multiple comparison corrective techniques is not found in the reporting and use of educational statistics such as those from the PISA.

That the field of educational assessment does not address multiple comparisons in a similar way to the scientifically oriented pharmaceutical sector provides another point of comparison between the practices of educational assessment and that of empirical science. It is also a distinction observed by Meehl (1978), who considers fields such as school psychology as “scientifically unimpressive and technologically worthless” (p.806). Meehl (1978) goes on to describe several difficulties in fields of psychology, or what he considers “soft areas of psychology”, and proposes as a solution that

Wherever possible, two or more nonredundant estimates of the same theoretical quantity should be made,... (p. 829)

This process is sometimes referred to as *converging operations* or *triangulation* (Garner, Hake, & Eriksen, 1956; Nunnally & Bernstein, 1994, p. 37), a test that shows a consistency of results and outcomes across two or more distinct trials. Meehl (1978) anticipates critiques of his proposed requirement for triangulation in areas such as educational assessment when triangulation is not used in other sciences. Meehl (1978) observes that

Consistency tests are so much a part of standard scientific method in the developed disciplines, taken so much for granted by everybody who researches in chemistry or physics or astronomy or molecular biology or genetics, that these scientists do not even bother having a special name for them! (p. 829)

Meehl (1978) goes on to suggest that consistency tests for “soft psychology” may be difficult to obtain. In the context of the broader discussion on educational assessment validity as science (Borsboom et al., 2003), Meehl (1978) is here suggesting that the character of educational assessment data is distinct from data generated by sciences with real entities such as physics and pharmaceuticals. This provides further evidence, as argued in Chapter 7, that the use of educational assessment is distinct from the operations of science.

One example where results from the PISA have been triangulated is in an analysis comparing the PISA and TIMSS by Wu (2009). Klieme (2016) also compares outcomes from the two programs. Wu’s (2009) analysis not only explores the numerical data, but

also the linguistically expressed research methodology and data on test material to reconcile the results from both programs. The analysis of Wu (2009) and Klieme (2016) describe similarities between programs as well as reconcile differences by drawing on broader evidence including the linguistic components of each assessment program.

Meehl (1978) presents further argument that educational sciences are distinct from the natural sciences. Nevertheless, this argument is distinct from the argument regarding the separation of science and ethics in Chapter 7. That is, while programs such as the PISA might not be able to communicate scientific truths, quality programs like the PISA can nevertheless provide strong warrants to support claims based on its data (Kane, 2006; Toulmin, 1958/2003). However, these claims remain open to rebuttal and might require further support, in a similar fashion to how Moss (2016) describes actual interpretation. Approaches to synthesising the scientific, ethical, and sociopolitical in educational assessment is now explored through the PISA.

8.2. The PISA and normative validity

Biesta (2015b, pp. 348-349), among others, considers the PISA as the most visible and among the most influential global educational measurement systems (Bonal & Tarabini, 2013; Breakspear, 2012; Grek, 2009). This section uses the PISA as an example of a competent subject in contemporary educational assessment. The documents produced by the PISA, particularly the PISA Technical Standards (OECD, 2017f), will be considered here as an explication of a contemporary approach to educational assessment that seeks to synthesise normative and technical validity.

The PISA Technical Standards (OECD, 2017f) focus on the implementation of the PISA and its data collection. In the context of the three tiers of educational assessment described in Chapter 3, and illustrated in Figure 2, the PISA Technical Standards (OECD, 2017f) belong in Tier-2 that describes the planning and implementation of an educational assessment. Tier-2 is empirically interrogable through documents such as the PISA assessment frameworks (OECD, 2017b), as well as the PISA Technical Report (OECD, 2017e), and related quality assurance data. The framework for interrogating Tier-2 is through pretheoretical concepts made evident in Tier-1 and which is the focus of this study. This analysis of the PISA is therefore consistent with explicating the knowledge of the competent subject as described in Chapter 3.

The PISA national involvement standard

The PISA Technical Standards (OECD, 2017f) are organised into three broad types of standards. The first are *Data Standards* that directly concern the scientific quality of the collected data. The second are *Management Standards* that outline responsibilities for timelines and effective project coordination. The third, and the departure point for this discussion, are the *National Involvement Standards* which address the sociocultural and sociopolitical dimension of the PISA.

The National Involvement Standards are designed to ensure that the assessment instruments used in the PISA “achieve cross-national, cross-cultural and cross-linguistic validity” (OECD, 2017f, p. 451). These involvement standards require participating countries to develop mechanisms to promote participation in the design of the PISA and to promote the dissemination of results. Quality assurance measures for this standard include participating countries documenting strategies for garnering feedback, as well as maintaining lists of committees, membership records of committees and groups, and relevant meeting records. Of interest to the present discussion is that the involvement standards require participating countries to provide feedback “on the development of instruments, domain frameworks, the adaptation of instruments, and other domain-related matters” (OECD, 2017f, p. 451).

While this study does not examine the efficacy of the procedures used by the PISA, nor the quality of involvement by participants, of interest are three principles exemplified by the involvement standards. First is that the PISA considers itself an activity oriented towards cultural validity and cross-national validity. Second is that the PISA considers the development of the domains and constructs assessed as being informed by feedback from participating countries. That is, the PISA considers the domain and construct assessed, or the latent variable, as somewhat human constructions. Third and finally, the PISA considers that feedback on domain-related matters from participating countries needs to be provided by properly constituted groups and committees keeping appropriate records. Two central conclusions can be drawn once these three principles are considered typical and exemplary of contemporary educational assessment practice.

One conclusion, which reinforces an observation already made in this study, is that the PISA is an activity that works in a different and irreconcilable paradigm to the one

advocated by Borsboom et al. (2003). Where Borsboom et al. (2003) advocate for entity realism where latent traits are considered to exist independent of human construction and explication, the PISA demonstrates that what it measures is a human construction that is politically and culturally dependent. Further, the paradigm of Borsboom and Wijsen (2016) considers that

[p]olitics does not mandate methods of scientific research or standards of justification; science is responsible for this by its own. (p. 282)

The paradigm of Borsboom and Wijsen (2016) is inconsistent with the political nature of the PISA which has a governing board responsible for setting priorities and standards (OECD, 2017e, p. 26). The PISA also places an onus on participating countries to provide feedback, which is also contrary to the paradigm described by Borsboom and Wijsen (2016) that seeks to separate politics from assessment. Finally, the argument for validity put forward by Borsboom et al. (2004) – which requires latent traits to exist as attributes and be causally responsible for measurement outcomes – is unsustainable in the PISA context where each domain is demonstrably constructed by humans through a focus on cross-cultural validity. The PISA provides a clear case that the entity realism advocated by Borsboom et al. (2003) does not apply to educational assessment, not even as a borderline case.

A second conclusion is that the involvement standards are consistent with Habermas' (1992/1998) concept of deliberative democracy described in Chapter 5. The involvement standards seek to institutionalise, through committees and groups that are well-formed in terms of composition and record keeping, a cycle of justification and application (Habermas, 1994). While the efficacy of the deliberative procedures for the PISA cannot be determined by this study, the documented involvement standards provide a clear illustration of principles underpinning contemporary educational assessment design and implementation. The diminishing capacity for national involvement in progressive cycles of the PISA is addressed later in this section.

Secondary analysis and actual interpretations

The model of intended and actual interpretation in educational assessment argued for by Moss (2016) is also evident in the PISA, which provides opportunity for secondary analyses of its data. In particular, the PISA provides opportunities for secondary analyses

that focus on the policy contexts of individual countries (OECD, 2005a, 2009b). Through the provision of advice, algorithms, and methods, the OECD encourages the broader international research community to explore the PISA data to enhance meaning in interpretation. The PISA claims that it “has identified hundreds of relationships between students, schools and, to a lesser extent, country characteristics and performance” (OECD, 2009b, p. 26). In providing this advice, the OECD (2009b) emphasises that the “structure of educational systems” (pp. 26-28) affects the collection of the data, data definitions, and survey outcomes. The OECD encourages secondary analyses not only to bring greater clarity to the relationships between the variables it has identified, but also to enhance the local actual interpretations for enhanced meaning relevant to educational structures within participating countries.

The facilitation of secondary analyses by the PISA corresponds to the research agenda proposed by Moss (2016) that seeks to explore how “professionals in different roles and contexts interact with and use data” (p. 248). It also corresponds to Habermas’ (2012a; 1994, 2003) reciprocal discourses of justification and application. The dialectical nature of the PISA explored here reinforces the argument that educational assessment data is a generalised form of communication where application, meaning and validity needs to be linguistically redeemed for each normative lifeworld context. While this study identifies processes and support for actual interpretations in local contexts by the PISA through secondary analyses, the efficacy of these processes requires empirical verification.

Inherent legitimation tensions of the PISA

The PISA illustrates an inherent legitimacy tension in terms described by Habermas (1975/2005). One proposed explanation given by Habermas (1975/2005) for legitimacy crises is the “unintended side effects (politicization) of administrative interventions in the cultural tradition” (p. 50). In the context of the PISA, as the PISA seeks to systemically integrate educational meaning across national boundaries, there is a tendency towards adverse and unintended side effects on the cultural traditions in those countries, such as the PISA measures not reflecting activity in participating economies (Ananiadou & Claro, 2009; Gorur & Wu, 2015). Habermas (1975/2005) conceptualises these crisis tendencies from the perspectives of system integration and social integration. On the one hand, citizens seek to broaden lifeworld horizons (Habermas, 2001, p. 83), a desire the PISA

responds to by providing a symbolic media of system integration for education systems across the globe. On the other hand, the level of system integration provided by the PISA becomes differentiated from lifeworld contexts to reach into the lifeworld to adversely affect lifeworld reproduction (Habermas, 1981/1992). This crisis tendency arising from tensions between system integration and social integration identified by Habermas (1975/2005) is reflected in the PISA.

There is growing evidence of unease towards the PISA and how, in Habermas' (1981/1992) terms, the systems nature of the PISA overpowers the lifeworld. This involves a decoupling of the numerical component of educational assessment from the linguistic component. Biesta (2015b) points to, and is critical of, the seductive nature of the PISA, with the PISA providing policy makers with "clear, unambiguous and easy to digest and to communicate information about the apparent quality of educational systems" (p. 350). Biesta (2015b) argues that this easy to digest information, particularly quantitative information, can overshadow narrative accounts and qualitative analyses of educational quality at the school and system level. Similarly, Gorur and Wu (2015), in the context of Australian public policy, point to how policy ambitions based on global system rankings in the PISA can mask the nuanced "differences between what is valued in Australian curricula and what is tested in PISA" (p. 662). These reflect broader concerns among the international academic community about the effects of the PISA on school systems, with its focus on quantitative measures, short-term policy cycles, and a relatively narrow range of measurable aspects (Ball, 2015; Grek, 2009; Lingard, 2011; "OECD and PISA tests are damaging education worldwide – academics," 2014). These tensions can be couched in terms of a legitimation crisis that arises from tensions between system and lifeworld.

Eroding validity of the PISA and the reporting of trends

The extensive critiques of the PISA suggest that its legitimacy is eroding, and the reporting of trends provides one possible explanation for this erosion (Biesta, 2015a; Bonal & Tarabini, 2013; Breakspear, 2012; Gebhardt & Adams, 2007; Grek, 2009; "OECD and PISA tests are damaging education worldwide – academics," 2014). Trends address the longitudinal progress of countries on the PISA scale over successive cycles, and in a way that implicitly reifies what is considered valued educational content and in

a way that eschews progress narratives. While this reification assists in the creation of numerical trend data that feed into economic models and planning (G. S. Becker, 1964/1993; Friedman, 1953/2008, 1969/2008, 1962/2008; Hanushek, 2016; Hanushek & Ettema, 2017), trends also lead to educational content being considered static and reified. It is through the reporting of trends that the PISA is sheared from the dialectic with technological and social progress in societies. The reporting of trends provides one possible explanation for the eroding legitimacy of the PISA.

The technical process of equating and trend reporting in educational assessment is elaborate (Dorans et al., 2011; Gebhardt & Adams, 2007; Holland & Dorans, 2006), however the fundamental principles are relatively straightforward. Holland and Dorans (2006, p. 194), after a review of the linking and equating literature, identify five requirements for valid equating and linking. One of the conditions is that the constructs of two assessments must be equal for valid equating and linking, and hence for trend reporting. That is, two tests need to measure the same thing for them to be equated for the purpose of trends (Dorans et al., 2011, p. 23). The PISA reports trends from the year 2000 to 2015 (OECD, 2016b, p. 78; 2017e), and the requirements for equating and linking (Dorans et al., 2011; Holland & Dorans, 2006) suggests that the nature of the test content remains the same over successive cycles. The requirement for construct equivalence identified by Holland and Dorans (2006, p. 194) might hold conceptually, literally, or both for the PISA. That is, while the nature of the test content might change in response to technology, the conceptual requirement for construct equivalence in trend reporting renders any discourse around changes in test content silent. It is the technical requirement for construct equivalence over successive cycles that is of interest to this discussion.

How equating and linking can be affected by changes in construct definition is illustrated by OECD's (2010b) *Computer-Based Assessment of Science* in 2006 in which three countries participated. This assessment involved students being assessed on both paper-based and computer-based forms of the test. The construct explications for both tests was ostensibly the same, with the computer-based version of the test consisting of items specifically designed for administration through a computer to include videos, simulations and animations (OECD, 2010b, pp. 12, 17). However, quite distinct results were obtained between the paper-based and computer-based modes of the assessment, with some items showing distinct gender differences that were difficult to explain

(OECD, 2010b, pp. 31-33, 85). The two modes were found to be comparable in terms of illustrating a difference between the two nodes, but incompatible for the purpose of equating due to each mode providing a distinct scale. That is, the way students ranked on the paper-based test was too distinct to how they were ranked on the computer-based test (OECD, 2010b, pp. 30,90). In a more contemporary example, a similar lack of comparability in computer-based and paper-based assessments was identified as an issue for NAPLAN (ACARA, 2018; Merlino, 2018; Robinson, 2018).

These instances further reinforce claims that constructs are made manifest in the test material (McCurry, 2017; Wiliam, 2010). That is, it reinforces that constructs manifest in the test in a way that is distinct from explicated definitions published in assessment frameworks and distinct from any conceptualised latent variable that transcends the test material (OECD, 2006; Rasch, 1960/1980; Wright, 1980). It also illustrates how the introduction of technology, even when the underlying explicated construct definition is the same, can affect the assessed construct and the process of linking and trend reporting.

The reporting of trends in the PISA, and the construct equality constraints of equating, largely keep conceptions of educational content static and hampers the PISA to embrace progress narratives. From both a technical and intuitive perspective, it is somewhat of a contradiction to report trends while updating assessment frameworks. When reporting trends, the only new content permitted is that which does not adversely affect the scaling of the original items. For example, the 2006 computer-based version of the PISA test described above was incompatible as the construct changed with the introduction of technology (OECD, 2010b). The reporting of trends is antagonistic towards the inclusion of educational assessment data that emerges from technological and social developments in societies and is consistent with the neoliberal economics of Friedman (1962/2008) that considers wants as fixed. This contradiction of implicitly keeping content definitions the same while societies progress technologically, leads to contradictory logics in schools to potentially affect developments in curricula. As described in the literature review in Chapter 2, these contradictory logics are demonstrated by the studies of Solomon and Lewin (2016) as well as Hardy and Lewis (2016). This study finds that the legitimisation tension of contradictory logics might be attributed to, in part, the reporting of trends in large-scale assessment.

While the initial legitimacy of the PISA might have resulted from effective processes of justification and application (OECD, 2017f, p. 451), trend reporting diminishes the need for this involvement. That is, once the initial construct definition has been established, the reporting of trends limits the effect that country feedback can have on evolving construct definitions, with selection of new items for subsequent cycles being constrained by the need to develop stable links. The PISA 2015 technical documentation states that the “more link items with good fit across groups, the more stable the link becomes” (OECD, 2017e, p. 143). That is, new items need to psychometrically fit existing items to make the link between cycles psychometrically valid and stable. The technical requirements for linking, along with the limited release of test materials (OECD, 2017e, p. 54), present possible causes for the erosion of the legitimacy of the PISA over time.

As a final comment, the analysis in this study suggests that some form of trend reporting is desirable for purposes of system integration at the intergovernmental and jurisdictional level, with the main issue emerging from trends resulting from *how* they are reported. For example, the *consumer price index* continues to be a useful economic indicator in light of regular debate about the relevance of the “basket of goods used” (Hausman, 2003). However, economic indicators tend to be reported with less illocutionary force (Austin, 1962/1975) than the PISA results. The PISA has the power to create the identity of countries, for example Finland has had its educational identity formed somewhat through the PISA (Breakspear, 2012; Chung, 2010; Grek, 2009). While trend reporting may be useful at the government level, it can have maladaptive effects when actual interpretations are made inappropriately at the local level and at the student level. This study observes that these effects might be ameliorated if trends were reported with a reduced illocutionary force.

Australia’s NAPLAN equates and reports trends in a similar fashion to the PISA across cycles (ACARA, 2015a). The reporting of trends provides a further explanation for the motivating issues for this study related to a malaise in educational assessment and Australia’s decline on the PISA scales (Masters, 2013; Thomson et al., 2013; Thomson et al., 2017). The reporting and use of trends cause a bifurcation between the knowledge and skills required in developing economies and societies, and how knowledge and skills are measured in accountability regimes such as the PISA and NAPLAN. This bifurcation

resulting from the reporting of trends provides a possible explanation for a loss of legitimacy, and a loss of coordination capacity for the respective assessment regimes.

8.3. Legitimation tensions emerging from progress narratives

This study has explored tensions in educational assessment emerging from different attitudes towards progress narratives. Further, it has worked towards a theory to address these tensions through Kane's (2006) approach to validation extended through Habermas' (1981/1985, 1981/1992, 1983/1996, 1992/1998) sociology and Toulmin's (1958/2003) use of argument. This section enumerates some of the identified legitimation tensions which are illustrated in Figure 13 and further described below. Completeness is not claimed for this list, with further legitimation likely to be identified through further discourse. However, Figure 13 provides broad response to the study question that addresses how different attitudes towards progress affect educational assessment design and use.

pragmatism	←	ideology	→	neoliberalism
wants as fluid	←	wants and demands	→	wants as fixed
communicative action	←	system integration	→	marketing action
administrative management	←	system management	→	strategic management
justification and application	←	citizen participation	→	stakeholder management
constructivist stance	←	ontology for trait	→	realist stance
lifeworld	←	primary perspective	→	system
ethical, rational	←	preferred reasoning	→	scientific, empirical
linguistic	←	symbolic domain	→	numerical
domain sampling <i>constellation metaphor</i>	←	evidence model	→	continuum <i>ruler metaphor</i>
performative attitude	←	student performance	→	performativity
intersubjectivity	←	operational focus	→	machine efficiency
for communication	←	Illocutionary force	→	for influence
actual interpretation	←	interpretation focus	→	intended interpretation

Figure 13 – Legitimation tensions in educational assessment

Ideology

Ideology describes a set of beliefs and convictions that bind groups together (I. Buchanan, 2010, p. 243). This study has adopted the ideological positions of pragmatism in the tradition of Habermas (1998) and in a fashion where it is in dialectic with both neoliberalism and postmodernism (Hayek, 1944/2007; Popper, 1945/2002; Stedman Jones, 2012). Triandis (2001), in a review of the literature on the personalities of collectivism and individualism found that people “sample from both individualist and collectivist cognitive structures, depending on the situation” (p. 909). The collectivism and individualism divide is also a formulation used by Hayek (1944/2007) in a fashion analogous to the pragmatism and neoliberalism divide used here (Hayek, 1944/2007, p. 64). The work of Triandis (2001) therefore suggests that the ideological dispositions of

pragmatism and neoliberalism are fluid among individuals and groups, and don't necessarily reflect the reified views of social actors.

The terms pragmatism and neoliberalism are used here to provide a contemporary overarching ideological framing for the tensions identified through this study. As described in Chapter 1, this study primarily seeks to distinguish each ideology's attitude towards progress narratives. Pragmatism is characterised as addressing issues emerging from progress through structures involving discourse ethics, deliberative democracy, and processes of justification and application (Habermas, 1994, 1983/1996, 1992/1998, 2003; Piaget, 1968/1973; Weber, 1946/2009a). In contrast, neoliberalism is characterised as seeking to address progress through markets with minimal government involvement (Friedman, 1962/2002; Hayek, 1944/2007; Kotler, 1972; Popper, 1945/2002; Smith, 1776/1993). This remainder of this section presents legitimation tensions in educational assessment in terms of attitudes that broadly correlate to the dispositions of pragmatism and neoliberalism.

Wants and demands

The pragmatic tradition is characterised as engaging with progress narratives to consider wants as fluid, dynamic and malleable to education (Bourdieu, 1979/1984; Meyer, 2000; Peters, 1959/1973). In contrast, the neoliberal tradition considers wants as largely fixed (Friedman, 1962/2008). When wants are considered fluid, an integrative system is required to coordinate the relationship between educational assessment and developments in society. This study observes that the wants as fixed assumption best matches models used by monetarist economics that are associated with neoliberal philosophies (Friedman, 1962/2002, 1953/2008, 1969/2008, 1962/2008).

System integration

Educational assessment has a role in system integration which addresses how autonomous systems of society support social integration that in turn responds to social differentiation emerging from progress narratives (Mouzelis, 1992, 1997). Habermas (1981/1985) addresses system integration pragmatically with a focus on communicative action. In contrast, neoliberalism addresses system integration through markets which, using the definition of marketing by Kotler (1972), can be cast as a form of strategic action in Habermas' (1981/1985) tradition.

Where the pragmatic tradition has a primary focus on communicative action for system integration, the neoliberal tradition seeks to integrate systems through markets. Tension around the ATAR, a symbolic medium for coordinating student transition from secondary to tertiary education in Australia, was used to explore this tension in Chapter 6 (Blyth, 2014; Finkel, 2018; Matters, 2015). The pragmatic tradition through its universal focus supports symbolic media such as the ATAR as it promotes social differentiation and social integration while maintaining universal enfranchisement. The universal approach of the ATAR allows students to pursue a wide range of studies in state-based local jurisdictions, while maintaining enfranchisement to a national universal system for accessing tertiary education.

The neoliberal tradition in preferring market-based methods of system integration tends towards proprietary university entrance tests targeting specific careers such as medicine. These market-based methods are perceived as being better at addressing instrumental concerns such as predictive validity (B. Griffin et al., 2018; Mercer et al., 2018; Poole et al., 2012).

The pragmatic tradition prefers that educational assessment programs used to coordinate activity in society reflect societal values established through deliberative democracy. Pragmatism seeks to promote social cohesion in contrast to the neoliberal tradition which is more comfortable with market-based educational assessment to coordinate system integration, and considers social fragmentation resulting from markets unproblematic.

System management

The pragmatic tradition prefers administrative forms of system management that focus on unity of purposes and bureaucratic process for managing inputs (Fayol, 1949/2013; Weber, 1946/2009b). Administrative management better supports system integration through formal procedures and law and is better placed to promote postconventional reasoning (Kohlberg, 1971). Administrative management focusing on pragmatism is also better at managing processes of justification and application (Habermas, 1994), and better at developing mechanisms for universal inclusion (Habermas, 1983/1996).

In contrast, the neoliberal tradition is characterised as preferring strategic forms of management that allow operations to be decoupled from central planning through numerical media (Capon, 1996; Jelinek, 1979; Mintzberg, 1994). Strategic management

allows systems across society to be more autonomous and decoupled from each other to promote both competition and flexibility. Strategic management relieves the burden on moral reasoning at the postconventional level (Kohlberg, 1971). Strategic management of systems also tend to reify conceptions of content and somewhat relies on a realist stance towards the latent variable for communicating expectations.

Citizen participation

The pragmatic tradition is characterised as considering educational assessment as setting norms and standards for society that are dynamic and evolving (Mislevy, Steinberg, & Almond, 2003). The pragmatism justifies norms and standards communicated through educational assessment and seeks to be responsive to feedback resulting from their application (Habermas, 1994, 2003). In contrast, the neoliberal tradition considers collective citizen participation as potentially endangering freedom and leading to indoctrination (Popper, 1945/2002, p. 106). However, in doing so the neoliberal tradition has difficulty in developing and justifying norms that are implied in terms such as “approved” as used by Friedman (1962/2002). The neoliberal tradition communicates with society as a market through stakeholder management rather than meaningful deliberative engagement (Donaldson & Preston, 1995; Eskerod & Huemann, 2013).

Ontology of traits

Following Borsboom et al. (2003, 2004), the pragmatic tradition is associated with a constructivist stance towards the latent variable assessed by educational assessment. This follows from the assumption that wants are fluid in a way that affects broader dynamics in the economy and society at large. Pragmatism considers educational constructs contestable and requiring ongoing processes of justification and application through ordinary language. A pragmatic perspective therefore argues for administrative forms of system management to coordinate these developments across the different spheres of society.

In contrast, the neoliberal tradition is characterised as preferring a realist stance towards the constructs assessed by educational assessment. Friedman’s (1962/2008) assumption that wants are fixed is consistent with the realist stance that considers the latent variable to exist independently of explication (Borsboom et al., 2003, 2004). Both the wants as fixed assumption and the realist stance diminish the requirement to explicate the latent

variable of educational assessment and avoids the need to pragmatically negotiate over educational constructs. The realist stance allows autonomous systems in society to communicate using the numerical component of educational assessment, without the need to extensively refer to its linguistic component.

Primary perspective

The pragmatic tradition through its focus on ordinary language has primary focus the lifeworld of education. The lifeworld perspective considers the world as experienced by students and teachers as active participants. In this way, the pragmatism is concerned with the effect that systems have on students and teachers in the classroom where lifeworld horizons are limited. In contrast, the neoliberal tradition is characterised as primarily taking a system perspective. In this way, neoliberalism focuses on steering media, such as money and its correlates that are encoded with meaning for coordinating societal activity.

As a corollary, the pragmatic tradition has a primary focus on educational assessment that is normatively anchored in classroom and school contexts. In contrast, the neoliberal perspective has a primary focus on institutionally anchored educational assessment with ever broadening horizons such as NAPLAN and the PISA.

Preferred reasoning

Habermas (1981/1992, p. 280) distinguishes between ties that are empirically motivated through inducement and deterrence, from ties motivated rationally through trust. This study characterises pragmatism as preferring a focus on reasoning that is rationally motivated, based on trust, and which has an ethical focus. This attitude provides for the accommodation of a plurality of views and perspectives. This study also identifies that this form of consensus becomes more difficult with broadening horizons. In contrast, the neoliberal tradition is characterised as preferring empirically motivated ties that allow actors to maximise goal-orientations for efficiency and profit.

Symbolic domain

Pragmatism takes ordinary language as its primary symbolic medium of coordination in contrast to neoliberalism which prefers numerical media. In this way, pragmatists focus on meaning-making in the lifeworld of children and students, while neoliberalism focuses on broadening horizons through numerically-based steering media. This tension between

symbolic domains is the source of tension between the system and lifeworld identified by Habermas (1981/1992), as well as the process of colonisation where coordination involves numerical media becoming decoupled from its linguistic meaning.

Evidence model

The literature review in Chapter 2 described how the domain sampling model, or constellation metaphor, is distinct from the continuum model, or ruler metaphor (Mislevy, Steinberg, & Almond, 2003; Reckase, 2017; Wright, 1997). The domain sampling model and constellation metaphor provides for a greater granularity of information about how students relate to a construct and to determine differences between students. The constellation metaphor is therefore preferred by pragmatists as it can report at a more granular level on what students know and can do. Enhanced granularity in reporting better supports and facilitates student individuation.

In contrast, the continuum model is better for ranking and comparing students, and more amenable to generalising across lifeworld horizons through linking and equating (Dorans et al., 2011; Holland & Dorans, 2006). The continuum model is therefore preferred by neoliberal approaches to education and is more amenable for use in markets and for empirically motivated strategic action.

Student performance

This study has identified the performative attitude as described by Habermas (1981/1985) and Weir (1995) as contrasting to performativity as described by Lyotard (1979/1984). The pragmatic tradition has a focus on individuation and socialisation to therefore place greater value on the performative attitude of students. The performative attitude allows students to create and express their own identities in cognitive, moral, and aesthetic dimensions. In contrast, the neoliberal tradition is characterised as considering wants and desires as largely fixed. Following this, the neoliberal tradition is characterised as preferring performance assessments that lead to performativity and which are more amenable towards efficiency measures and marketisation.

Operational focus

Given that the pragmatic tradition considers the constructs of educational assessment fluid and dynamic, pragmatism has a focus on intersubjectivity. To this end, pragmatism prefers human involvement in test development, item-writing, invigilation, and marking

of student responses. In contrast, the neoliberal tradition, through a concern for efficiency prefers machines and algorithms to replace teachers to achieve instrumental efficiency.

Illocutionary force

When educational assessment is conceptualised as a generalised form of communication, actual interpretation of educational assessment data requires validity to be redeemed in reference to the local context (Moss, 2016). To this end, pragmatism prefers the illocutionary force of educational assessment reporting to be consistent with its communicative aims, which is generally to simply describe or constatae a student's performance on a test. In contrast, the neoliberal tradition seeks to increase the illocutionary force of reporting to enhance its capacity to steer activity through markets and through creating and conferring identity (Marginson, 1993, 1997b; Mockler, 2013). Further, this study identifies that the excessive illocutionary force through which NAPLAN results are reported in Australia, provides one possible explanation for the maladaptive student responses identified in the study by Howell (2012, 2017) (Lipnevich et al., 2016).

Interpretation focus

Following Moss (2016), this study has cast the performance-sanction paradigm of Parsons and Smelser (1956/2005) as the reporting of an intended interpretation by one system and an actual interpretation made in another. Following Habermas (1994), this study has also cast the intended interpretation in terms of reporting universal norms that need to be justified, and actual interpretation in terms of application that makes reference to local contexts.

This study identifies the pragmatic tradition as placing greater emphasis on application in actual interpretation, and the neoliberal tradition emphasising the reporting of universal norms in intended interpretation, particularly as a form of market pressure.

Further, this study notes that this dichotomy has an association with gender. As described in Chapter 2, Gilligan (1977, p. 486) takes a contrasting approach to Kohlberg (1971) on moral development to argue that women are generally more sensitive to context and conflict in making moral judgements. In this sense, it could be argued that women might place greater importance on actual interpretation than the intended interpretation. While no essentialist claims on gender are required in this context (Diquinzio, 1993), the work

of Gilligan (1977) does alert to the potential for individual differences in attitudes towards universal norms and their concrete application, and that these differences may not be solely due to ideology.

Postconventional reasoning

While this study is in the pragmatic tradition, the legitimation tensions enumerated in Figure 13 are not presented by way of showing how pragmatism is opposed to neoliberalism. Instead, and consistent with the pragmatic tradition (Beiser, 2005; Bowie, 2003; Hegel, 1807/1977), this study is in dialogue with neoliberalism. What the legitimation tensions enumerated in Figure 13 asks of neoliberalism is that in concrete situations and where relevant and necessary, each tension be resolved through recourse to the better argument drawing on relevant information. The legitimation tensions are therefore presented to facilitate discourse over educational assessment so that it can better meet the needs of society and its citizens, and in a way that leads to prosperous economies.

This study characterises pragmatism as responding to the immediate normative context of the lifeworld of students, with neoliberalism seeking to broaden those normative horizons. In this sense, both ideologies respond to distinct needs of citizens that are inherently in conflict. This study therefore argues that the design and use of educational assessment, and notions of validity, involve judgements along the dichotomies described in Figure 13, where application depends on judgements made in concrete situations, contexts, and scenarios. These judgements may need to trade between circumscribing life-project opportunities through a focus on local contexts and extending student lifeworld horizons through symbolic media that may alienate them (Ball, 2016b; Habermas, 1981/1992). Judgements may also need to be made that balance the human need for intersubjectivity (Hegel, 1807/1977) with the management of scarce resources through machine efficiency (Bower & Christensen, 1995; Hammer, 1990; Hammer & Champy, 1994). These judgements are all potential subjects of substantial argument in educational assessment validation (Habermas, 1983/1996; Kane, 1990, 1992, 2006; Toulmin, 1958/2003).

The work of Habermas (1983/1996, 1975/2005) suggests that the inherent crisis tendencies in educational assessment can only be ameliorated through postconventional reasoning. That is, in terms of the stages of moral development described by Kohlberg

(1971), conventional moral reasoning is grounded in an immediate lifeworld that only requires loyalty and conformity to a shared normative background. When educational assessment is considered from a conventional level of moral reasoning (Kohlberg, 1971), attempts to move beyond the normative horizons of individual actors results in a loss of coordinating capacity. These tensions can be reconciled through postconventional reasoning, and a failure to reconcile them through postconventional reasoning provides a possible explanation for the malaise in educational assessment described by Masters (2013).

Postmodern rejoinder

The legitimation tensions described in Figure 13 are characterised in terms of tensions between dichotomous positions towards which postmodernism takes a different stance. The tensions are considered by this study as inherent in educational assessment design and use in societies advancing in technological and social dimensions with their resolution involving recourse to moral development (Kohlberg, 1971). Derrida (1972/1981), in the postmodern tradition, seeks to avoid such binary positions, where “[o]ne of the two terms governs the other” (p. 41) through a violent hierarchy. Derrida (1972/1981) goes on to argue that such binary positions should be deconstructed, and that binary oppositions “inhabit philosophical opposition, resisting and disorganizing it, *without ever* constituting a third term, without ever leaving room for a solution in the form of speculative dialectics.” (p. 43). This is to suggest that postmodernism does not seek to address the inherent legitimation tensions in educational assessment identified here, and instead seeks to deconstruct in search of a third position.

Derrida’s (1972/1981) approach to deconstructing binary oppositions to create third positions is in the realm of speculative research of the subject and not in the intersubjective realm of fields like educational assessment (Rorty, 1995). The argument by Derrida (1972/1981) is consistent with the position of Ball (2013), who proclaims that “I am more interested in the crafting of an academic subject yet to come” (p. 21). While philosophical speculation may indeed lead to new third positions, it also leads to somewhat of a performative contradiction (Watt, 1975) when these matters are addressed in concrete situations. The speculation of both Derrida (1972/1981) and of Ball (2013)

are out of scope of this study's question that seeks to address the more immediate needs of citizens in contemporary societies.

Closing remarks

This chapter has built on previous chapters and worked towards a reconstruction of educational assessment and towards a general theory of educational assessment validity. It has argued that education address three broad purposes, that of qualification, socialisation, and the self. Further, following the argument that educational assessment is an inherent part of education, it has argued that educational assessment addresses these three broad purposes, and may or may not generate persistent data.

This chapter has made the case for an increased focus on normative validity in educational assessment (Biesta, 2009). Educational assessment communicates societal values and expectations (Mislevy, Steinberg, & Almond, 2003) which give rise to legitimation tensions that require recourse to discourses of justification (Habermas, 1994; Rehg, 2011). A number of other normative issues exist in education, such as the tension between permanence and transience of interpretations (Messick, 1989; Moss, 2003), to pose the question as to when it might be appropriate to reify a student's identity through reporting an educational assessment result. The importance of semiotic resources as part of normative validity in technology-based contexts is another issue raised in this chapter (Kress, 2003, 2010). Further, the use of multiple comparisons in the interpretation of large-scale educational assessment data provides a scientific justification for triangulation when making actual interpretations in normative contexts. These analyses all illustrate how educational assessment can become decoupled from its normative anchor to compromise normative validity.

This chapter has also described how the PISA Technical Standards (OECD, 2017f), and particularly the National Involvement Standards, work towards the process of justification and application for the PISA. The processes documented in the PISA standards provide an explanation for the legitimacy garnered by the PISA (Biesta, 2015b; Bonal & Tarabini, 2013; Breakspear, 2012; Grek, 2009). This chapter has also speculated how the reporting of trends and equating, which reifies constructs and conceptions of knowledge, may be eroding the contemporary legitimacy of the PISA (Dorans et al., 2011; Gebhardt & Adams, 2007; Holland & Dorans, 2006; "OECD and PISA tests are

damaging education worldwide – academics," 2014). The chapter closed by identifying a number of inherent legitimation crises for educational assessment, and further highlighted the role of postconventional reasoning to address issues emerging from progress narratives in society.

Chapter 9. Discussion

This chapter reflects on the broader argument and case made in this study. The first section reinforces how different ideological approaches seek to address issues emerging from progress narratives in different ways. The pragmatic tradition is characterised as being able to reconcile these competing ideological positions through recourse to the better argument institutionally established through deliberative democracy. The second section revisits the key themes and issues for educational assessment identified in this study, mainly contests around the conceptualisation of the construct, the emergence of new mathematical models, how the imperative for normative validity affects test development, and the role of reporting and its effect on stakeholders. The third and final section revisits the issues and academic imperatives motivating this study.

9.1. How ought educational assessment be designed

This study's response to the question on how different attitudes towards progress narratives affect educational assessment design and use is broadly summarised in relation to the legitimisation tensions outlined in Figure 13. Through identifying the tensions, this study concludes that educational assessment design be subject to the better argument in concrete situations through deliberative democracy that includes discourses of justification and application (Habermas, 1994, 1992/1998). Consideration to these processes in the context of educational assessment are addressed here.

This study has framed issues in educational assessment around legitimisation tensions that arise from progress narratives which generate complexity towards which different ideologies take distinct approaches. Pragmatism is characterised as preferring to talk through these tensions, neoliberalism as preferring markets to address them, and postmodernism as turning away from them altogether. This study has also identified a tension between notions of intersubjectivity and notions of the individual and the subject (Habermas, 1981, 1997, 2001). Where pragmatism works towards universal values of the group that are intersubjectively determined, neoliberalism argues on behalf of individual freedoms, and postmodernism advocates on behalf of the subject (Ball, 2016b; Hayek, 1944/2007; Popper, 1945/2002; Wittig, 1993). This study has developed a case that the different ideological perspectives can be reconciled through substantial argument that is

sensitive to the needs of all citizens in society (Habermas, 1992/1998; Toulmin, 1958/2003).

Pragmatism— laws and norms

This study has developed its case from the pragmatic tradition of Habermas (1998) and education's relationship to law provides a compelling reason why the pragmatism should be the leading orientation in educational assessment. The primacy of Habermas' (1998) pragmatic approach emerges from education being a universal activity that is legislatively mandated for citizens in all OECD countries (Kant, 1803/2009; OECD, 2016a, p. 475; Plato, 1963). It is through the compulsion of law applying to student participation in education that generates an imperative for educational assessment to meet the needs of all citizens in society.

Pragmatism has a well-established academic literature towards the functional and structural aspects of society particularly through the work of Parsons (1963a), Parsons and Smelser (1956/2005), and Habermas (1992/1998). Nevertheless, this study has found education and educational assessment backgrounded by these traditions in a way that is incommensurate with education's role in knowledge-based economies (Drucker, 1968/1992). Habermas (1992/1998) relegates education to the cultural sphere of society and as part of the lifeworld, and similarly Parsons and Smelser (1956/2005) consider education part of the latent pattern maintenance system of society.

The contemporary importance of education and educational assessment is illustrated through the growing and continued influence of human capital theory in policy development (G. S. Becker, 1964/1993; Finkel, 2018; Hanushek, 2013, 2016; Marginson, 1989, 1993, 1997b). In considering education a background cultural activity, the Habermas' tradition has somewhat vacated the field in terms of education's relationship to the economy and polity more broadly. This retreat has left the field open to neoliberal economic traditions such as that of Hanushek (2013) and Hanushek and Ettema (2017), who respectively conceptualise productivity and efficiency in terms of numerical ratios without reference to content.

The retreat by pragmatism has also left the field open to postmodern critique particularly critique that builds on the work of Lyotard (1979/1984) such as that of Ball (2003, 2013, 2016b). This study has sought to address this gap in the pragmatic tradition to foreground

educational assessment as a structural component of knowledge-based economies (Drucker, 1968/1992).

In foregrounding educational assessment as symbolic media and a vital structural component for knowledge-based economies (Drucker, 1968/1992), this study has sought to give educational assessment a similar status to law. This status arises from education largely being a sphere of society constituted in law through mandated schooling. In Australia, for example, all formal qualifications are legislatively integrated through the Australian Qualifications Framework (AQFC, 2013). In Victoria, the *Education and Training Reform Act 2006* mandates compulsory education, establishes a formal role for public schools, and specifies the role and responsibilities of the teaching service. Furthermore, the act is responsible for establishing a curriculum and assessment authority, a teaching institute for maintaining the registration of teachers, and an authority for maintaining a register of approved educational institutions. The act also legislates the role of education in relation to vocational education and training, higher education, and adult education. The clear case of education being foregrounded in law is at variance with Habermas' (1992/1998, pp. 196, 360) conceptualisation of education as a cultural activity where the law is only foregrounded in cases of conflict (Murphy & Fleming, 2010, p. 3). This study has sought to foreground educational assessment in relation to other spheres of society such as the economy and polity, and to enhance the conceptualisation of the relationship between educational assessment and society's laws and norms.

By foregrounding the relationship between norms expressed through educational assessment and laws expressed through legislation, this study elevates the role of item developer and test developer to one comparable with that of legislator. Where legislators establish laws, test developers in educational assessment establish norms (Habermas, 1992/1998; Rawls, 1971/1999). In a similar fashion to how law makers legitimate their deliberations through discourses of justification (Habermas, 1994, 2003), agencies responsible for generating educational assessment symbolic media might similarly justify their work to enhance its legitimacy.

The role of educational assessment agencies is also elevated in knowledge-based economies where human capital theory drives public policy (G. S. Becker, 1964/1993; Finkel, 2018; Hanushek, 2013, 2016; Marginson, 1989, 1993, 1997b; Rudd & Gillard,

2008). It is through human capital theory that the symbolic media generated by educational assessment is combined with the steering media of money through mixed ratios. In this sense, educational assessment agencies deserve a similar status, and to be subject to similar scrutiny, to central banks such as the *Reserve Bank of Australia* (Reserve Bank Act 1959) and the *Federal Reserve System* in the United States (Federal Reserve Board, 2005). This study finds the legislative frameworks that currently apply to educational assessment underdeveloped. The level of disinterest observed in the regulation of central banks is not similarly observed in educational assessment agencies where strategic private interests and markets have significant influence (Eacott, 2017; Hogan, 2012, 2016a, 2016b; Hogan et al., 2015).

It is in the relationship that educational assessment has, through human capital theory, with measures of money that an enhanced role for ethics in large-scale educational assessment emerges. While this study seeks to foreground education and distinguish education from the cultural system of society, education nevertheless remains an essential part of the cultural system. Where, in Habermas' (1981/1992) terms, money is largely a delinguistified steering media, educational assessment measures remain connected to normatively and culturally anchored meaning that is linguistically mediated. The onus is on the pragmatic tradition to better conceptualise the relationship that educational assessment has with the cultural sphere of society and with the economy through the concept of ethics.

There is a neoliberal concern that state-controlled education endangers freedom and can lead to indoctrination. This concern is explicitly expressed by Popper (1945/2002, p. 106) and this study finds that it emerges from educational assessment's intractable connection to language. It is a concern that can be expressed in terms of moral development described by Habermas (1983/1996). When educational assessment symbolic media is generated from a conventional level of moral reasoning (Kohlberg, 1971), the norms expressed through it reflect and favour a particular sociocultural point of view. This can be perceived as indoctrination by spheres of society not sharing those norms. Habermas (1983/1996) seeks to ameliorate these effects through postconventional moral development appropriate for pluralistic societies. In this sense, neoliberal concerns over indoctrination are somewhat justified, but nevertheless are concerns that should be addressed through the pragmatic tradition. In particular, in pluralistic societies there is greater onus on

educational assessment test developers to adopt a postconventional moral attitude and to justify the norms that their tests promulgate (Habermas, 1994; Kohlberg, 1971).

This study identifies three key challenges for the pragmatic tradition in educational assessment. One is that deliberations over educational assessment can become bogged down and cumbersome. This linguistic burden is evident in the account by Barcan (1996, 2003) of educational reform in Australia during the 1980s. A second concern is that sections of society may be disregarded in deliberations over educational assessment. This study has addressed this concern in relation to gender, with Fensham (2016) illustrating how the subject of physics can be affected through how the perspective of girls are incorporated into the curriculum and assessment (Hildebrand, 1996). While this study does not seek to arbitrate on these issues, it does identify that pragmatism enters a performative contradiction when its claim to universality involves ignoring certain sectors of society. These two challenges directly relate to the third challenge, which is to consistently develop appropriate levels of postconventional reasoning and inclusion in educational assessment design and use (Habermas, 1983/1996; Kohlberg, 1971). To better address these challenges, this study advocates the adoption of administrative forms of system management that are more amenable to developing appropriate rules of bureaucracy for inclusion (Fayol, 1949/2013; Weber, 1946/2009a).

Neoliberalism – empiricism and efficiency

The appetite for empiricism and optimisation within the neoliberal tradition is evident in the work of Hayek (1988/1980, 1944/2007) and Friedman (1962/2002, 1969/2008, 1962/2008). Friedman (1962/2008, pp. 12-13), after three perfunctory paragraphs addressing the fluidity of wants, nevertheless takes as a point of departure the assumption that wants are fixed to mathematically model the economy in terms of demand. A similar haste towards mathematical optimisation is evident when Friedman (1962/2002) addresses education using the term “approved” in relation to schools and schooling. There is a distinct avoidance of rational debate on matters of plurality and a divergence of views over meaning, an avoidance explicitly expressed by Caldwell and Hayward (1998) in the case of Victorian education policy in the 1990s who considered progress forces such as feminism a challenge to authority. Instead, neoliberalism is associated with a distinct

sense of individualism and goal-oriented social action towards maximising return from human and physical capital (Friedman, 1969/2008; Hayek, 1944/2007).

This study remains open to the possibility that the market-based monetarist economics of Friedman (1953/2008, 1969/2008, 1962/2008) have efficacy in militating against economic crises and system crises in general (Habermas, 1975/2005). Further, market-based models open possibilities for new intersubjectively shared lifeworlds consistent with the impulse of citizens to broaden lifeworld horizons (Habermas, 2001, p. 83). Nevertheless, this study identifies that monetarist and market-based models tend to assume wants as fixed so that the rules of the game can be similarly fixed to allow individuals to maximise opportunities. Neoliberalism eschews time-consuming processes associated with deliberative and collective processes that interfere with perceptions of efficiency. However, a market-based approach as advocated by neoliberalism enters a performative contradiction when it justifies and bases its argument on the concept of personal freedom (Friedman, 1962/2002; Hayek, 1988/1980, 1944/2007) when participation in education is compelled by the state.

There is an imperative for neoliberalism to engage in pragmatic discourse on what is valued and what is legitimate in education and educational assessment (Biesta, 2009; Habermas, 1975/2005). This imperative emerges from education being universally legislatively mandated (OECD, 2016a) and occurs in the realm of citizenship and not in the realm of markets. Chapter 6 identified a distinction between a social actor being in a market and a social actor being a citizen. Markets are defined in terms of social actors with an interest and the resources to engage in an activity (Kotler, 1972), and citizens are social actors bound together through law (Habermas, 1992/1998; Rawls, 1971/1999). It is through education's relationship to law that an imperative emerges for neoliberalism to engage with pragmatists in good on what is valued in education. For the compulsory years of schooling, education is a public good which is legitimated through the body politic and not through markets.

Postmodernism – an escape from what is

This study has cast postmodernism as addressing those excluded from societal discourse and those for whom society is an imposition. Postmodernism addresses the concerns of the subaltern who is excluded from participation in society by the cultural hegemony of

elites (Gramsci, 1971/2014; Spivak, 1985/2010). Postmodernism also addresses the oppression of classes unable to express themselves as a subject under Marxism (Braaten, 1995; Wittig, 1993), and those with libidinal energies that exclude participation in mainstream frameworks and economies (Bowie, 2003; Foucault, 1984/1988, 1976/1990, 1984/1990; Lyotard, 1979/1984, 1974/2015). The paralogical reasoning and the little narratives proffered by Lyotard (1979/1984), and the tactical appropriation of time proffered by de Certeau (1984), provide a sense of temporal relief from oppressive structures. Nevertheless, this study considers this relief temporary with the onus on pragmatists to ensure long-term universal inclusion (Habermas, 1981/1992).

While there are justified claims that structuralism and pragmatism can lead to repression, particularly claims from the feminist tradition (Fraser, 1990, 2013; Ingram, 2010; Meehan, 1995), postmodernism is also associated with an escape from tradition. Rorty (1995) describes what he calls radicals, and what Habermas (1981, 1997) calls conservatives, and this study calls postmodernists, as seeking to break out “of some particular inheritance (a vocabulary, a tradition, a style) which one had feared might bound one’s entire life.” (p. 457). Rorty (1995) goes on to associate this quest for escape with a search for the ineffable and the sublime, and as a quest that is inherently a private concern, and a quest that focuses on the subject at the expense of intersubjectivity. This study has identified this attitude in the work of Ball (2003, 2013, 2016b), and identified it as an approach dominant in contemporary educational critique (Gillies, 2013). Habermas (1981, p. 13; 2001, p. 147) describes postmodernism as presaged on concepts from modernism which it then appropriates in its critique of modernism. This study finds this a performative contradiction and argues that there is an onus on the postmodern tradition to meaningfully reengage with structuralist approaches of society on matters relating to educational assessment.

Synthesising perspectives – the integrative system

This study has identified a role for an integrative system in society through the work of Parsons and Smelser (1956/2005). The integrative system is envisaged as synthesising the ideological perspectives used in this study. In the context of the inherent legitimisation crises identified by Habermas (1992/1998, 1975/2005), the integrative system is also

envisaged as mediating and harmonising how dynamics in the economy and other systems of society affect educational assessment design and use.

Parsons and Smelser (1956/2005, p. 57) do not propose concrete organisational units for their AGIL scheme. Habermas (1992/1998) similarly avoids specifying organisational units when elaborating deliberative democracy for managing legitimisation tensions. Instead, Habermas (1992/1998, p. 367) identifies organisations emerging spontaneously from the private sphere as well as an organised public sphere for garnering legitimacy. In a similar fashion, this study does not nominate discrete organisational units for the integrative system.

The integrative system therefore includes institutions across society such as institutions for system integration responsible for generating educational assessment symbolic media, as well as institutions of social integration such as schools (see Chapter 5, Figure 5). In the main, organisations responsible for generating educational assessment symbolic media are responsible for the intended interpretation described by Moss (2016), and for discourses of justification described by Habermas (1994). Similarly, organisations that directly deal with citizens and students are responsible for actual interpretation as described by Moss (2016), and for leading discourses of application described by Habermas (1994). There may also be intermediary institutions, such as central offices in jurisdictional education systems, which may be responsible for translating international research results for local contexts. In this case, central offices would be responsible for both actual interpretations and translating them into new intended reinterpretations appropriate for local contexts. This study therefore considers the process of validation and legitimisation a responsibility dispersed across organisations in society.

By anchoring educational assessment validation through principles of deliberative democracy, the processes of validation and legitimisation becomes part of public policy formulation and consistent with its principles (Althaus, Bridgman, & Davis, 1998/2013; Bardach, 2012; Dawes, 2010; Moore, 1995, 2013). Locating the processes of validation and legitimisation into public policy does not supersede the approaches to educational assessment validity and validation as described by Cronbach and Meehl (1955), Messick (1989, 1994), and Kane (1992, 2006, 2016a), among others. Instead, it calls on the processes of validation to be incorporated into organisations and institutions that have a

role in public policy development. In both metaphorical and conceptual terms, deliberative democracy emerges from this study as a container for educational assessment validation (Bion, 1963).

The requirement for educational assessment design to articulate with broader public policy is not a novel suggestion and is instead consistent with contemporary practice. This study has identified ongoing developments in the relationship between educational assessment and law, particularly in the United States on matters of civil rights (Gordon Commission, 2013; S. E. Phillips & Camara, 2006). What this study proposes is that the connection between educational assessment and other spheres of society be strengthened. *The Standards* (AERA/APA/NCME, 2014) have been identified as addressing ethical and legal issues, particularly those that arose from the civil rights movement in the United States. However, this study also identifies that *the Standards* (AERA/APA/NCME, 2014) might not be able to inform the development of educational assessment as symbolic media that address broader economic and cultural concerns. This claim is supported by the PISA developing its own set of technical standards (OECD, 2017f) that address the relationship educational assessment has with the economy and culture. One lever for ensuring educational assessment meets the needs to economies, societies and citizens is therefore further work on standards addressing educational assessment design and use.

An imperative for further work on standards governing the design and use of educational assessment is illustrated by developments associated with NAPLAN in Australia. The PISA developed its own set of technical standards independent of *The Standards* (AERA/APA/NCME, 2014) to achieve cross-national and cross-cultural validity (OECD, 2017f, p. 451). However, where the PISA reports at the jurisdictional level, NAPLAN reports at the student level and uses procedures that “are similar, if not identical, to those used in the Programme for International Student Assessment” (ACARA, 2015a, p. 2). This is to suggest that NAPLAN may be working towards standards implied by the PISA standards and in a way that allows it to work outside of requirements related to matters such as opportunity to learn as stipulated in *the Standards* (AERA/APA/NCME, 2014). The PISA reports at the jurisdictional level, while NAPLAN reports at a student level for which the PISA standards were not designed.

9.2. Implications for educational assessment

This study has made a case for expanding the scope of concerns addressed through educational assessment validation to include those related to legitimacy. Several important contemporary policy issues in educational assessment design and use also emerge from this expanded framing. The policy issues are discussed here in relation to the construct, evidence, development and reporting nodes of the Koomen and Zoanetti (2018) framework.

Construct – realism, constructivism, operationalism

The construct node has been identified as an area of considerable contestation in the educational assessment validity debate (Borsboom et al., 2003, 2004; Crocker, 1997a; Newton & Baird, 2016a; Shepard, 2016). This contestation emerges from educational assessment addressing an indivisible property of both society and the student, and that educational assessment refers to a dynamic relationship that students have with the economy and polity. Following Wu et al. (2017), constructs can be considered as being somewhat fluid and shaped by context and purposes of educational assessment. These characteristics of educational assessment are incompatible with the demands of validity made by Borsboom et al. (2003, 2004) for example.

Nevertheless, this study also observes approaches to educational assessment that take a realist stance towards the latent variable, approaches that seek to find causal factors within students responsible for educational assessment outcomes. For example, Didau (2016), following Shakeshaft et al. (2013), argues that assessment outcomes can be explained by DNA, and argues that society can have differential expectations of some students for this reason. The realist stance is also evident in certain forms of education system management that rely on expectations and results being communicated numerically without reference to the linguistic component of educational assessment. The realist stance removes the burden on communicative action and the burden to communicate substantive meaning across contexts. This study finds that the realist stance in educational assessment diminishes its coordinating capacity through its loss of connection to language and meaning.

There is support for McCurry (2017) and Wiliam (2010), who both describe the construct as becoming manifest in the content of an assessment. However, the idea that the

construct remains bound to test content is incompatible with contemporary approaches to linking and equating that require, or assume, constructs across different tests to be considered equal across different tests (Dorans et al., 2011; Holland & Dorans, 2006). The argument that constructs become manifest in the test content is also incompatible with assumptions made in item response theory to challenge the use of that theory (Hambleton et al., 1991; Masters, 1982; Rasch, 1960/1980; Yen & Fitzpatrick, 2006). The transition to NAPLAN online (ACARA, 2018; Merlino, 2018; Robinson, 2018) and the computer-based assessment of science by the PISA (OECD, 2010b) support the claim that the introduction of technology changes the nature of constructs as they are made manifest in a test, even if the assessment framework informing item development remains the same. The computer-based assessment of science in the PISA (OECD, 2010b) and the transition to NAPLAN online, support the proposition that the nature of educational measures evolve in response to technological development in society (Papert, 1993; Vygotsky, 1978). In doing so, these studies support the argument that constructs become manifest in the test material to challenge the fundamental basis and justification for equating, linking and trend reporting (Dorans et al., 2011; Gebhardt & Adams, 2007; Holland & Dorans, 2006).

Evidence – constellation or ruler metaphor

The educational assessment paradigm that builds on the domain sampling model and the constellation metaphor has the potential to provide more granular information about the knowledge, skills and abilities of students (Mislevy, Steinberg, & Almond, 2003; Reckase, 2017). There is therefore a tension between the domain sampling paradigm and the continuum paradigm currently dominant in Australian education and closely associated with item response theory (Adams et al., 1997; Andrich, 1978; Hambleton et al., 1991; Masters, 2013; Reckase, 2017; Wright, 1997). Item response theory is attractive for use in human capital theory as it provides measures that can be used at the ratio level that are amenable to mixed ratios with units of money (G. S. Becker, 1964/1993; Fischer, 1995; Hanushek, 2016; Hanushek & Ettema, 2017). These calculations become more complicated, if not impossible, in the domain sampling model or constellation metaphor. A recent Gonski et al. (2018) report recommended the development of an on-demand student learning tool for the purpose of formative assessment through which teachers

would be able to tailor teaching to meet the individual needs of students. The constellation metaphor provides the better model for implementing such a system. However, as noted by Reckase (2017, p. 8), test development processes for domain sampling are distinct from those used in the continuum model. Domain sampling requires an explication of a domain and for each part of a domain to be represented in an educational assessment. In contrast, item development for the continuum model proceeds on the basis of desirable psychometric properties that reference a hypothesised unidimensional latent variable (Hambleton et al., 1991; Wu et al., 2017). Furthermore, the constellation metaphor is associated with mathematical models such as Bayesian models (Almond et al., 2015; Mislevy, Steinberg, & Almond, 2003) that are distinct from models associated with item response theory (Adams et al., 1997; Adams & Wu, 2007; Yen & Fitzpatrick, 2006).

In Chapter 1, an argument was made that Australia's leadership in education and educational assessment resulted from leadership in item development (McCurry, 2017) and psychometrics (Adams et al., 1997; Andrich, 1978; Masters, 1982). A change to a constellation metaphor or a domain sampling model might better suit the purposes of on-demand assessment in the fashion recommended by a recent Gonski et al. (2018) report. However, a change to the constellation metaphor would require a major cultural shift for the Australian educational assessment sector currently tied to the ruler metaphor.

Development – normative validity

This study has identified test development as important for normative validity. Mislevy, Steinberg, and Almond (2003, p. 4) identify a role for educational assessment in communicating expectations and norms which is a claim supported by practices around setting performance standards (Baird et al., 2000; Cizek, 2012b; Cizek & Bunch, 2007; Hambleton & Pitoniak, 2006). Further, this study finds that the practice of establishing norms involves discourses of justification (Habermas, 1994, 2003). Discourses of justification are evident, for example, in the implementation of the VCE in Victoria where the VCAA make publicly available: study designs that describe content on exams; past exam papers; and examiner's reports (www.vcaa.vic.edu.au). However, this study also finds normative validity, and discourses of justification, as coming under increasing pressure (Biesta, 2015a).

The broadening horizons of educational assessment and the reporting of trends are two developments this study identifies as challenging notions of normative validity. Through broadening the horizons of an educational assessment, the cultural specificity and curricular validity of test material can diminish. Grisay (2002) illustrates this potential when describing the “dropping” of test material that does not perform uniformly across countries in the PISA. McCurry (2017) also illustrates this when describing stimulus-based cross-curricula test items for testing across jurisdictions, a form of testing also adopted by the PISA (OECD, 2000). These processes can diminish the normative validity of test content.

The reporting of trends can also diminish the normative validity of test content. First, the reporting of trends reifies notions of what is being assessed through the requirement for construct equivalence (Dorans et al., 2011; Holland & Dorans, 2006). Furthermore, the reporting of trends requires link items to be kept secure over successive cycles for programs such as the PISA. This results in a limited release of test materials to dilute and diminish discourses of justification and application (Habermas, 1994, 2003; OECD, 2017e, p. 54). This poses an ongoing challenge to the normative validity of large-scale assessment programs.

This study identifies normative validity as an important matter as it directly pertains to the evidence used in evidence-based decision-making. Following Gorur and Wu (2015), policy development based on numerical rankings is in the realm of analytic argument and not substantial argument. Gorur and Wu (2015) put the case that arguments that arise from educational assessment should be based on the content of educational assessment. This argument also pertains to evidence used in the application of human capital theory (G. S. Becker, 1964/1993). Blaug (1976) and Tan (2014) both identify studies of human capital theory that produce empirical evidence in discord with the theory. This study suggests the discord between empirical evidence and theory may be the result of the linguistic component of the educational assessment data used in the theory not appropriately reflecting activity in the economy. Following Blaug (1976) and Tan (2014), this study does not argue for the abandonment of human capital theory, and instead argues for a greater consideration of the linguistic component of educational assessment, for which normative validity is important.

Finally, this study identifies the need for further work on how normative validity is affected by enhanced semiotic resources afforded by technology (Kress, 2003, 2010). This study finds a continued focus on the migratory strategy identified by Binkley et al. (2012) in technology-based educational assessment, and not on the transformational strategy. The contemporary focus on the migratory strategy is evident in the reporting of trends in programs such as NAPLAN and the PISA (Dorans et al., 2011; Gebhardt & Adams, 2007; Holland & Dorans, 2006). The reporting of trends, and the requirement for construct equivalence, leads to a focus on comparability between modes at the expense of exploiting technology's transformational potential in meaning-making (Beavis, 2010; Kress, 2003, 2010; OECD, 2010b, 2017e; Robinson, 2018). There is a role and responsibility for educational assessment to reflect, establish, and reinforce the normative validity of the semiotic resources used in domains across society.

Reporting – performativity and steering media

This study has developed the notion of educational assessment as symbolic media for coordinating activity in society. Following Moss (2016), reporting is considered as involving intended interpretations with actual interpretations made in local contexts. Furthermore, actual interpretations may be made in systems outside of education such as in the economy on matters related to employment. When making actual interpretations, access to institutional evidence may be diminished, and access to normatively anchored additional evidence enhanced. Through conceptualising educational assessment as symbolic media, this study has identified a need to develop further principles and standards for reporting educational assessment results.

Parsons (1963a) identifies requirements for institutionally anchored symbolic media that can be applied to educational assessment reporting. The reporting of results needs to have a defined category of value or a value principle, a category of interest, a definition of the situation where the category applies, and a normative framework for using the educational assessment data (Chernilo, 2002). For example, Wu (2016), the American Educational Research Association (2015), and Harris and Herrington (2015), all argue that in certain circumstances it can be inappropriate to use educational assessment data to make inferences about students, teachers, or schools. Further, Tierney and Koch (2016) identify a need for privacy to assist in student learning. The development of standards addressing

the reporting of educational assessment data covering issues related to appropriate use and privacy would ameliorate these concerns (Berson & Berson, 2006; R. Buchanan et al., 2018). The precepts for symbolic media developed by Parsons (1963a) provide a good basis for developing reporting standards (Chernilo, 2002).

The work of Butler (1988, 1990/2007) suggests that education can construct gender identity, and this study extends this argument to suggest that educational assessment constructs student identity along cognitive, moral, and aesthetic dimensions. This is an issue explored by Reay and Wiliam (1999), who identify that students have different dispositions on how they react to feedback from educational assessment, and that for some students their test results contribute to their understanding of themselves. Lipnevich et al. (2016) identify both adaptive and maladaptive responses to feedback, and Howell (2012, 2017) presents empirical evidence to back this claim. Greater attention to the method of reporting, particularly its illocutionary force (Austin, 1962/1975), might ameliorate the effects of inappropriate identity construction and other maladaptive responses by students resulting from educational assessment.

Finally, this study has shown that educational assessment is not capable of communicating scientific truths. Unlike, for example, the pharmaceutical industry (Shi et al., 2012), the use of large-scale educational assessment data does not take into account issues of significance arising from multiple comparisons (Collis & Rosenblood, 1985; Davies & Goldsmith, 1988). Following Meehl (1978), this study finds educational assessment as scientifically unimpressive, and advocates for two or more nonredundant estimates for decision-making. This process is otherwise known as convergent operations or triangulation (Garner et al., 1956; Nunnally & Bernstein, 1994, p. 37). This is to suggest that the validity of actual interpretation in local contexts is not simply enhanced by further local evidence, but that its validity is dependent on such evidence as educational assessment data alone, which is not scientifically robust (Moss, 2016; Wu, 2014).

The argument that the validity of actual interpretation relies on local evidence because educational assessment is unable to communicate scientific truths provides further support to the argument that science and ethics be considered in a unitary fashion. That programs such as the PISA and the TIMSS cannot communicate scientific truths does not

mean that these programs do not provide the best available evidence. These programs can provide strong warrants to support claims based on their data (Kane, 2006; Toulmin, 1958/2003). However, both the ethics of discourse (Habermas, 1983/1996) and the science of multiple comparisons (Collis & Rosenblood, 1985; Davies & Goldsmith, 1988) provide imperatives to consider further evidence when claims are made in local contexts.

Moss (2016) argues for greater support for actual interpretations within local contexts and this study supports this claim. The PISA provides support through guidance for secondary analyses (OECD, 2005a, 2009b), and the VCAA provides support through the publication of past exam papers and examiners reports (www.vcaa.vic.edu.au). These processes in contemporary educational assessment are consistent with approaches towards symbolic media (Chernilo, 2002; Parsons, 1963a, 1963b), and consistent with processes of justification and application (Habermas, 1994, 2003). In this way, validation is not simply an expert issue but an issue for all users of educational assessment data.

9.3. Looking ahead from the motivating issues

This section briefly revisits the issues motivating this study and gaps in the academic literature to which this study seeks to contribute. This studies focus has been on different attitudes to progress narratives. This focus has emerged from three somewhat disparate issues: a sense of malaise in educational assessment (Masters, 2013), issues with technology-based projects (IBAC, 2016, 2017; Pearson, 2012; Robinson, 2018), and Australia's performance on the PISA scales (Thomson et al., 2013; Thomson et al., 2017). Further, this study finds these issues somewhat unexpected as Australians led global developments in educational assessment programs such as the PISA (Adams & Wu, 2002; McCurry, 2017; OECD, 2000). Nevertheless, while these issues are somewhat surprising and disparate, one possible explanation arises around the notion of performativity which inherently eschews progress narratives (Ball, 2003; Lyotard, 1979/1984).

A disavowal of the progress narratives does not stop progress from occurring and does not stop the process of creative-destruction and disruption in modes of production (Fabo & Drahokoupil, 2016; Karimi & Walter, 2015; Schumpeter, 1942/2008). Instead, a failure to engage with the narrative of progress affects the development of symbolic media used to coordinate society. This study has identified the monetarist economics which models wants as fixed (Friedman, 1953/2008, 1969/2008, 1962/2008) and large-scale educational

assessment that models constructs as fixed (Dorans et al., 2011; Holland & Dorans, 2006; Rasch, 1960/1980), as failing to engage with progress narratives. In this way, the symbolic media used to coordinate society, such as money and educational assessment, all model activities in society in a way that assumes its basic structure and modes of production are fixed. This causes the meaning encoded in symbolic media to shear off from societies that continue to progress along technological and social dimensions.

While this study has not sought to detail the narrative of progress in the material structure of society (Marx, 2000; Schumpeter, 1942/2008), it has nevertheless taken it as given. In particular, this study has identified the discontinuity arising from the emergence of the knowledge-based economy described by Drucker (1968/1992). It has also drawn on the notion of business process reengineering of Hammer and Champy (1994), and of Hammer (1990), who not only argue for automation, but for the obliteration of certain activities in societies. The changing nature of production and demand is also explored by Jones (1982/1995) who identifies a relationship between the nature of demand and the structure of the economy. This study finds the material structure of society as changing rapidly through disruptive technologies (Fabo & Drahoukoupil, 2016; Karimi & Walter, 2015), even if the symbolic media used to coordinate society does not to-date reflect these changes.

The divergence between the material structure of society and how these are modelled and talked about manifest in various ways to generate legitimation crises. Legitimation crises emerge when symbolic media no longer accurately represent activity in the lifeworld (Habermas, 1981/1992; Husserl, 1954/1970). Habermas (1981/1992) casts these effects in terms of system imperatives bursting the capacity of the lifeworld to colonise it. Postmodernism casts this effect in terms of simulacra and simulation, where symbolic media no longer reflects any reality at all (Baudrillard, 1981/1994; Jameson, 1991). These are different ways of expressing the phenomena where the material and social reality of society is no longer reflected in the symbolic media used to coordinate it.

The loss of legitimacy in media is not an issue isolated to educational assessment. There is growing concern in general mainstream media with notions of “fake news” (Allcott & Gentzkow, 2017; Lischka, 2017). There is also scepticism towards media generated by scientific activities on matters related to climate science through notions of “climate

change denial” (Häkkinen & Akrami, 2014; Lewandowsky, Oreskes, Risbey, Newell, & Smithson, 2015). Where this study’s motivating questions address the legitimacy of educational symbolic media, failures in the legitimacy of symbolic media in general are a broader issue for societies across the globe. It is this slippage between activity in society and the symbolic media used to coordinate it that provides the basis for this study’s explanation for the contemporary malaise in educational assessment.

Australian education encased by data sets

Weber (1920/2011) used the term “steel-hard casing”² (p. 177) to describe the drive for empirical teleological efficiency for an industrial era, and this study suggests it as an appropriate metaphor for contemporary education. This study finds contemporary Australian education encased by data sets reflecting static notions of educational content bound together by equating methodologies that assume construct equivalence where content is considered unchanging across space and time (Dorans et al., 2011; Gebhardt & Adams, 2007; Holland & Dorans, 2006). This study explains its motivating issues in terms of a shift from linguistically focused communicative action that are able to incorporate progress to numerically focused strategic action through static data sets (Habermas, 1981/1992). The dominance of NAPLAN and the PISA lead to the characterisation of Australian educational discourse as encased and circumscribed by data sets.

Through explaining this study’s motivating issues in terms of a shift from communicative action to strategic empirically-motivated action, this study associates these forms of action with broader concepts through pattern language (Alexander et al., 1977; Gamma et al., 1995). Communicative action is here associated with substantial argument over what is true, right and beautiful (Toulmin, 1958/2003). These arguments are linguistically mediated and associated with rationalism and motivated by trust (Habermas, 1981/1992). Communicative action is also associated with notions of ethical discourse concerned for principles for universality and inclusion (Habermas, 1983/1996).

In contrast to communicative action, strategic action is associated with empirical and analytic arguments that are purposively motivated. In this way, strategic arguments are

² *stahlhartes Gehäuse*, translated as ‘steel-hard casing’ here, is notably translated as ‘iron cage’ in other translations. See Kalberg’s translation of Weber (1920/2011, p. 397) for discussion.

also associated with numerical steering through data banks that describe reified concepts of meaning. It is the reification of content and the empirical and strategic attitudes towards its use that frame this study's motivating issues towards which this study works towards a solution.

Malaise in educational assessment

Educational policy encased in reified data sets provides one explanation for the malaise described by Masters (2013). Lyotard (1979/1984, p. 51) describes data banks such as those of NAPLAN and the PISA as “nature” for the postmodern man that provide an endless basis for analytic arguments and insights. However, in Toulmin's (1958/2003) terms of analytic arguments, these insights are nothing more than restatements of the original data set itself. Jameson (1984) describes performativity as scientific work not concerned with modelling reality but instead concerned with simply producing new and fresh scientific statements. This sentiment is consistent with claims by the OECD (2009b) that the PISA data offers an “inexhaustible source of information for analysing educational issues” (p. 25). This is to suggest that educational policy is indeed encased in static sets of data.

Continued analyses of educational data sets raise the possibility of spurious conclusions when issues of statistical significance are ignored (Collis & Rosenblood, 1985; Davies & Goldsmith, 1988; Dunn, 1959, 1961; Stanovich, 1988). Further, consistent with the explication of performativity by Lyotard (1979/1984), these insights are based on a static view of society and can be exploited for commercial purposes with messages selected and distorted to meet the demands of certain audiences and market segments (Gillis et al., 2016; Kotler, 1972). Strategic and goal-oriented analyses of NAPLAN and the PISA data sets, targeting market segments, provides one explanation for the malaise and fragmentation identified by Masters (2013).

Habermas (1981/1992, p. 180) suggests that the coordination of societal action can only be transferred to symbolic media when strategic action is differentiated out. However, this study observes an abundance of strategically motivated and market-based action in educational assessment and education more generally (Burch, 2009; Burch & Good, 2014; Hogan, 2012, 2016a; Hogan et al., 2015, 2016; Kotler, 1972; Lingard et al., 2017; Marginson, 1997b).

The effect of markets in educational assessment as a symbolic media is an emerging issue. This study has explored educational assessment in relation to knowledge-based economies (Drucker, 1968/1992) and how educational assessment is used as symbolic media in a way comparable to that of money and administrative power (Habermas, 1992/1998). It also observes a proliferation of educational assessment as symbolic media, particularly in symbolic media coordinating student transition from secondary education to tertiary education (Edwards et al., 2012; Poole et al., 2012; VTAC, 2016). The proliferation of educational assessment generated symbolic media is analogous to the proliferation of crypto currencies for the media of money with Gandal and Halaburda (2014), for example, exploring the effects of competing crypto currencies (Blundell-Wignall, 2014; Brenig, Accorsi, & Müller, 2015). The proliferation of educational assessment as symbolic media is a threat to the legitimacy of the nation-state in a similar way to the proliferation of new economic media. Each symbolic media deserves similar attention as both affect social cohesion and the legitimacy of governments, particularly when they are combined through human capital theory (G. S. Becker, 1964/1993; Lyotard, 1979/1984).

Issues with educational assessment technology

An inordinate focus on instrumental and strategic action provides a further explanation for issues motivating this study. The traditions of Vygotsky (1978) and Papert (1993), for example, suggest that technological progress affects consciousness and affects meaning-making. These developments require engagement with new meaning through communicative action and are not simply matters of instrumental business efficiencies (Hammer, 1990; Hammer & Champy, 1994). These developments also place greater demands on social integration and system integration that require linguistic engagement and communicative action (Habermas, 1981/1992, 1975/2005). An inordinate focus on numerically-based instrumental and strategic action provide one overarching explanation for issues in implementing technological projects.

The failed Ultranet project, which was designed to provide an online learning and assessment platform, provides a poignant example (Pearson, 2012, p. 20). The subsequent official inquiries identified a persistent failure to create a coherent business case for the project, and a persistent lack of planning despite the involvement of a large number of

private consultancy firms (IBAC, 2016; Pearson, 2012). When reports from the inquiries are considered from the perspective of symbolic media (Chernilo, 2002; Habermas, 1981/1992; Parsons, 1963a), a question emerges over what the Ultranet was supposed to communicate. Furthermore, when the reports are considered from the perspective of intended and actual interpretations described by Moss (2016), a question emerges on how the actual interpretations of data communicated through the Ultranet to parents and teachers was supposed to be interpreted and used. There is no suggestion within the reports emerging from the inquiries that any consideration was given to the characteristics of educational assessment as a form of symbolic media requiring interpretation. The transcript of proceedings suggests only perfunctory recourse to academic expertise during the project through a report from the University of Melbourne which a witness describes as giving “us a lot of street credibility in terms of our capacity to tell other people” (IBAC, 2016, p. 381). The Ultranet project focused only on the instrumental aspects of communication to leave ethics and meaning-making unattended (Department of Education and Training Victoria, 2017).

A similar emphasis on strategic and empirically motivated action is evident in efforts to bring NAPLAN online. A. J. Martin and Lazendic (2018) frame the discourse of the troubled efforts to bring NAPLAN online through measurement precision and test efficiencies, and not through an interest in the enhanced communicative potential of technology-based educational assessment (Binkley et al., 2012; Kress, 2003, 2010; Merlino, 2017, 2018; Perelman, 2018; Robinson, 2018). Both the NAPLAN online endeavour and the Ultranet project suggest a lack of concern for communicative reason and meaning through an inordinate focus on efficiency and precision.

This study’s explanation for issues around the Ultranet project, and to a lesser degree NAPLAN online, is the general malaise in educational assessment identified by Masters (2013, p. 1). However, this study finds the unifying principle proffered by Masters (2013) itself a strategic and fragmenting intervention. Masters (2013, p. 12) argues that the fundamental purpose of educational assessment is to locate the position of a learner at the time of the assessment. Such a principle belies a preference for the ruler metaphor with which Masters (1982, 2013) himself is most familiar. Further, the principle suggests that once the location of a learner is established then that data or information can be communicated online through multiple channels for universal use. There is a sense the

Ultranet was envisaged on a premise consistent with Masters' (2013) principle. However, this principle ignores the requirement that each use of an assessment interpretation requires its own validity argument (Habermas, 1994; Moss, 2016; Newton, 2012). The failure to attend to this second interpretation in local contexts provides an explanation for loss of meaning in project coherence and hence overall project coordination. Greater attention to what is entailed in an actual interpretation might frame an alternative response for ensuring that educational assessment meets the needs of society and its citizens.

Masters' (2013) unifying principle suggests an instrumental focus through measures of "where" a learner is at the time of assessment (Masters, 1982; Wright, 1997). An alternative principle could be based around the constellation metaphor (Mislevy, Steinberg, & Almond, 2003) to focus on "who" a learner is at the time of assessment. A focus on "who" a student is provides greater scope for addressing the student-referenced and care-referenced concerns described by Cowie and Bell (1999) in approaches to classroom-based educational assessment. In seeking to provide a unifying mechanism for addressing the malaise, Masters' (2013) principle emphasises concerns for instrumental measures to neglect ethics and meaning making.

Motivating issues – Australia's rise and decline on the PISA scales

Australia's decline on the PISA scales (Thomson et al., 2013; Thomson et al., 2017), when it led the establishment of the PISA (Adams & Wu, 2002; OECD, 2000), is both a curiosity and a substantial question. This study provides an explanation for both its rise, its changes of fortune, and its decline.

This study explains the rise in Australia's educational fortunes through a focus on linguistic rationality and communicative action that arose after the second world war. The post-war period witnessed significant public investment in the education of Australian citizens. Public policy included a university scholarship scheme implemented by the Menzies government in the 1950s (Laming, 2012, p. 46). As the account by McCurry (2017) suggests, the "Baby Boomers' Scholarship Test" (p. i) of the 1960s led to capacity in system integration across Australian states and its education sectors. The scholarship program also led to the development of item-writing capacity that was later employed in the PISA. The post-war investment in education also fostered broader developments in education including developments in curriculum and system management (Barcan, 1996,

2003; Laming, 2012). One further outcome of the post-war investment in Australian education was the emergence of several Australian global leaders in numerically-based educational assessment methods based on the Rasch (1960/1980) model (Adams et al., 1997; Adams & Wu, 2002; Andrich, 1978; Masters, 1982; Wu et al., 1997). The post-war period is also associated with developments in communicative competency within Australia education, and with Keynesian (1953/1964) economics (Laming, 2012; Marginson, 1993).

This study identifies that the communicative capacity in Australian and Victorian education peaked in the late 1980s and early 1990s. This claim is supported by the career trajectories of Stacey (2001) and Fensham (2016) who both allude to an apex in Victorian curriculum development around the start of the 1990s (Blum, 1993; Stillman, 2007; Turner & Stacey, 2015). Furthermore, both Stacey (2001) and Fensham (2016) describe its subsequent decline throughout the 1990s. Nevertheless, both were able to personally move on from this period to contribute to the development of assessment frameworks for the PISA (OECD, 2000, p. 103; 2014, p. 465). These developments illustrate Australia's post-war capacity building in curriculum, assessment frameworks, and linguistic coordination of education, and its waning influence during the 1990s.

As described in Chapter 1, system coordination began to transfer to numerical media during the 1990s as Australia's psychometric capability emerged (Andrich, 1978, 1988; Masters, 1982; Wright & Masters, 1982). The LAP was introduced into Victoria during the 1990s and this assessment subsequently evolved into the national NAPLAN program that employs similar methods to the PISA (ACARA, 2015a; Pearson, 2009). As described, the requirement to report trends for both PISA and the NAPLAN leads educational content to be considered static through the requirement for construct equivalence for equating and linking (Dorans et al., 2011; Gebhardt & Adams, 2007; Holland & Dorans, 2006). In this sense, Australia's emerging capacity in numerically-based methods led to the imposition of a metaphorical steel cage around notions of educational content.

The emergence of Australia's capacity in numerical methods does not fully explain the transition to numerical steering in Australian education, as the post-war period was also associated with increasing contestation. Barcan (1996) identifies a range of interest

groups emerging in Australia from the 1970s including ethnic groups, environmentalists, and those focused on attitudes towards sexuality. Fensham (2016) also identifies a particular focus on feminist and environmental concerns emerging in Victorian education throughout the 1980s. The plurality of views in Australian education, along with the ongoing concern for education's relationship to economic growth, led to the 1980s being a period of considerable reform and contestation in Australian education (Barcan, 1996, 2003; Beare, 1983; Caldwell & Spinks, 1988; Finn, 1991; Marginson, 1989, 1993; Mayer, 1992). A transfer of steering from linguistic media towards delinguistified media is consistent with Habermas' (1981/1992) hypothesis of steering media emerging when the burden on ordinary language for coordination becomes overloaded. Further, Marginson (1997a) describes markets as providing an attractive option for governments to address this conflict, and numerical steering of educational outcome data provides one means for facilitating markets (Masters, 2005).

This study attributes the decline in Australia's educational achievement in terms of disengagement with substantial arguments on what is true, right and beautiful in contemporary education. This disengagement is attributed to the broader notion of performativity where the static data sets of NAPLAN and the PISA define what is valued in education to stifle linguistic discourse. These developments have led to social action in contemporary Australian education to be empirically motivated through artefacts such as numeric ranks and input-output ratios associated with performativity (Gorur & Wu, 2015; Hardy & Lewis, 2016; Keddie, 2016; Lambert et al., 2015; Solomon & Lewin, 2016). Australia's decline in educational achievement (Thomson et al., 2013; Thomson et al., 2017) can therefore be explained through a transfer from communicative action to strategic action.

Further empirical studies would need to verify the extent to which NAPLAN and the PISA have affected Australian educational discourse, however the studies described in the literature review in Chapter 2 are consistent with the explanation given here. The study by Howell (2012, 2017) on Australia's NAPLAN suggests that Australian students are more concerned about the empirical outcomes of assessment than the educational content on tests. The trend towards analytic argument is identified by Gorur and Wu (2015) who describe policy development based on empirically-based ranks, and not on substantive arguments about educational content. The trend towards empirically-based

performativity is also evident in the study by Lambert et al. (2015) that identifies a diminution of regard towards arts subjects in Australian education. These all support a claim that system coordination in Australia's education system has transferred from linguistic to numerical media, from rationally to empirically motivated action, from substantial to analytic argument, and from the performative attitude to performativity. These studies are all consistent with the hypothesis that NAPLAN and the PISA have created an empirical steel casing over Australian education.

The academic gaps

This study has responded to the opportunity provided by Flórez Petour (2015) who makes the case that educational assessment reform should be approached from historical, systemic and ideological perspectives. This study has demonstrated that these are fruitful lines of inquiry. Developments in educational assessment in both Victoria and Australia suggest that understanding the historical narrative is important to understanding contemporary issues. While the historical narrative is likely to be different for other jurisdictions, it is envisaged that some aspects of the narrative explored here are shared more broadly through the influence of global assessments such as the PISA.

The systemic approach advocated by Flórez Petour (2015) is fruitful as it provides a framing for exploring how system level educational assessments relate to those at the classroom level. This study has conceptualised this relationship through the perspectives of Parsons (1963a) and Habermas (1981/1992). However, this relationship can be conceptualised differently by others and therefore remains an open area of research.

The argument for considering ideological approaches to educational assessment made by Flórez Petour (2015) remains relevant. This study has identified ideological perspectives as affecting educational assessment design and use, and the need for these ideological perspectives to be reconciled due to the compulsory nature of education. This study has used the ideologies of pragmatism, neoliberalism and postmodernism to frame ideological perspectives. Other formulations and further perspectives may be more appropriate for other analyses. This study does not seek to reify ideologies but to stress the importance of the better argument in educational design and use where there are competing demands.

This study has extensively engaged with calls by Moss (2016) for a broader approach to validity that includes consideration of local users of educational assessment data. This call is consistent with the continued rise of knowledge-based economies (Drucker, 1968/1992). While this study has engaged with and sought to progress the agenda described by Moss (2016), there nevertheless remains considerable work on how intended interpretation relates to actual interpretation. Furthermore, the study by Howell (2012, 2017) suggests an imperative for further work on the effect of actual interpretation on students.

Baird and Lee-Kelley (2009) address the environmental complexity in the development of new educational assessments. Koomen and Zoanetti (2018) contribute to this agenda with an instrumentally focused framing, with this study seeking to provide an ethical framing that addresses how meaning and culture affects educational assessment design and use. Educational assessment design involves setting norms, expressing values, and communicating expectations (Baird et al., 2000; Mislevy, Steinberg, & Almond, 2003), and is therefore inherently a political activity. The political nature of educational assessment requires community consultation across broad areas of society through processes of justification and application (Arnstein, 1969; Habermas, 1994, 2003). In terms of the generic blueprints proposed by Baird and Lee-Kelley (2009), this study suggests that consultation around ethics and meaning may be more important than instrumental concerns around processes and technology in educational assessment design.

Finally, this study has responded to calls from Elwood (2013) for greater ethical considerations in large-scale assessment. The importance of ethics is a consistent theme in this study and follows Habermas (1976/1979) who reconstructs the work of Marx and Engels (1846/2000) to add moral-practical reasoning in the broader narrative of progress. The issue of ethics in educational assessment is a broad and persistent one and is an issue of intersubjectivity. The work of Howell (2012, 2017), that shows student maladaptive responses to feedback, is consistent with a lack of ethics in reporting educational assessment results to students. At the broader scale, the failure of the Ultranet project also demonstrates a lack of ethics in project management (Department of Education and Training Victoria, 2017). There is therefore an imperative for further research on how ethics apply in the design and use of educational assessment.

Closing remarks

This chapter has closed this study's argument on how different attitudes towards progress narratives affect educational assessment design and use. The study has sought to inform how contemporary educational assessments ought to be designed and has argued that educational assessment is a sociocultural activity that needs to be open to democratic influences. While this study argues that the approaches of both Messick (1989) and Kane (2006) are sound, they are also expert-based. In locating educational assessment in the broader sociology of Habermas (1976/1979, 1981/1985, 1992, 1981/1992, 1983/1996, 1992/1998), a framing is provided for how educational assessment is, and can be, influenced by democratic processes and recourse to the better argument.

When educational assessment is located within a broader sociological democratic context, the role of ideology in educational assessment become manifest. Ideology can influence how societies implement educational assessment, but there nevertheless remains an imperative for educational assessment design and use to remain open to the better argument. This imperative emerges from the compulsory nature of educational assessment that locates it as an activity of citizenship and not of markets.

This study has identified an expansion of themes addressed by educational assessment validation (Kane, 2006) to include concerns of legitimacy. It has also identified an expansion of roles and actors considered important in educational assessment validation. In orienting educational assessment towards both scientific and sociocultural concerns validation processes become more elaborate and contested in advancing economies. Enhanced processes of justification, application and validation in educational assessment through sound bureaucratic and administrative processes provides a way for managing these contests.

This study has responded to several issues identified in contemporary educational assessment implementation, as well as to gaps identified in the academic literature. A focus on numerical forms of coordination and management that emerges from the use of item response theory provides one explanation for this study's motivating issues. While item response theory provides a means for generalising results beyond their spatiotemporal context to assist with system integration, it can also lead to a lack of cultural precision and a reification of how knowledge and skills are conceptualised. An

inordinate focus on the numerical component of educational assessment fosters neglect in communicative action. A numerical focus, while assisting in system integration, also provides for markets which can relieve the burden for communicative action for system coordination. However, a shift away from communicative action leads to potential project implementation failures as well as a general decline in legitimacy of educational assessment among citizens in society.

Chapter 10. Conclusion

This study has located the field of educational assessment into the field of sociology to provide a broader framing for educational assessment design and use for contemporary knowledge-based economies (Drucker, 1968/1992). This has been achieved through three direct links. First, the approaches to argument by Toulmin (1958/2003) link the approaches to validation developed by Kane (2006) and evidentiary arguments in educational assessment developed by Mislevy, Steinberg, and Almond (2003) to the discourse ethics of Habermas (1983/1996). Through Toulmin (1958/2003) further links were made between educational assessment and discourse ethics (Habermas, 1983/1996), communicative action (Habermas, 1981/1985), deliberative democracy (Habermas, 1992/1998) and processes of justification and application (Habermas, 1994, 2003).

The second link between educational assessment and sociology made by this study is the casting of the intended interpretation and actual interpretation described by Moss (2016) into the performance-sanction paradigm elaborated by Parsons and Smelser (1956/2005). This conceptualisation provides for educational assessment to be considered as a form of symbolic media that allows for the coordination of autonomous spheres of society (Habermas, 1981/1992; Parsons, 1963a, 1963b).

The third link relates to the inordinate focus on the quantitative component of educational assessment identified by Wu (2014) among others (Gillis et al., 2016; Gorur & Wu, 2015; Wu, 2016). This study has articulated a distinction between the numerical component of educational assessment and its linguistic component. It is the process where these components become decoupled that coordination of activity in society is “transferred over to delinguistified media of communication” (Habermas, 1981/1992, p. 180). These three links provide a strong connection between the fields of educational assessment and sociology to better inform educational assessment design and use.

The reconstruction of educational assessment into discourse ethics exposes educational assessment to principles of universalisation (U) and principles of discourse (U) developed by Habermas (1983/1996). In this framing, educational assessment directly relates to the economy and polity in a way where it simultaneously addresses matters of scientific truth and ethical rightness. This study has identified an enduring relationship between the economy and educational assessment through human capital theory which links education

to economic growth (G. S. Becker, 1964/1993; Hanushek, 2013, 2016; Hanushek et al., 2013). Education's relationship to the economy provides one influence on educational assessment design and use. A second imperative emerges from broader political and legal considerations (AERA/APA/NCME, 2014; OECD, 2017f; S. E. Phillips & Camara, 2006), and considerations of fairness (Camilli, 2006; Zieky, 2016). In this way, this study's reconstruction of educational assessment through discourse ethics explicitly characterises educational assessment as a sociocultural activity.

In explicitly characterising educational assessment as a sociocultural activity, this study has engaged with contemporary arguments in assessment validity, particularly those made by Borsboom et al. (2003, 2004). Borsboom et al. (2003, 2004) advocate the realist stance, where validity is dependent on a latent variable existing and being causally responsible for observed student behaviour. In contrast, this study has made the case that the latent variable in contemporary educational assessment practice is a human construct that requires explication. It has been shown that the nature of constructs is often influenced by social, cultural and political considerations (OECD, 2017f). Nevertheless, this study has found that the approach advocated by Borsboom et al. (2003, 2004) applies to certain forms of psychological assessment, and therefore argues that validity progress along two distinct paradigms.

The validity paradigm emerging from Borsboom et al. (2003, 2004) might be applied to certain forms of psychological assessment that are able to support the realist stance and which are able to separate ethics from science (Borsboom & Wijsen, 2016). A second paradigm, explored through this study, can then emerge to specifically address educational assessment as a sociocultural activity oriented towards the economic, cultural and social spheres of society, and which engages science and ethics simultaneously (Messick, 1989; Shepard, 2016). This study identified overlap between educational assessment and some branches of psychological assessment. For example, the psychoanalytic (Freud, 1995), humanist (Rogers, 1959, 1951/2003), and evolutionary (Gebser, 1985; Kegan, 1982/1996, 1994/1997) approaches are also concerned with cultural factors in a similar fashion to educational assessment. However, behaviourist (Pavlov, 1927; Skinner, 1938, 1957) and biological approaches (Fernald, 2007) to psychology are consistent with the realist stance advocated by Borsboom et al. (2003, 2004).

An imperative for developing two paradigms emerges from a growing research interest in how biological factors affect behaviour (Habermas, 2003/2015), and how biological and genetic factors influence learning (Baron-Cohen et al., 2014; Didau, 2016; Kautz et al., 2014; Kovas et al., 2016; Shakeshaft et al., 2013). While this study does not discount the possibility that biological factors affect student performance in educational assessment, this study also identifies that contemporary practices in educational assessment do not seek to isolate these factors. Instead, contemporary practices in educational assessment address these matters in a way where ethics and science are simultaneously addressed through concepts such as fairness (Camilli, 2006; Masters, 2002; Zieky, 2016). Contemporary educational assessment is generally oriented towards the economy and polity in a manner indivisible from biological factors (OECD, 2017f; W. Schulz et al., 2016).

In developing the concept of educational assessment as symbolic media, this study has found that educational assessment is not able to communicate scientific truths. One reason emerges from programs such as the PISA not taking into consideration issues with multiple comparisons in a manner consistent with scientific principles (Collis & Rosenblood, 1985; Davies & Goldsmith, 1988; Dunn, 1959, 1961; Stanovich, 1988). This leads to a requirement for multiple nonredundant estimates (Meehl, 1978), converging operations (Garner et al., 1956), or triangulation (Nunnally & Bernstein, 1994), when making interpretations of educational assessment data. This position is consistent with Wu (2014) who argues that interpretation should be informed by personal experience and sense-making as well as numerical data. It is also consistent with Habermas (1994) who argues that discourses of application have their own arguments based on local contexts. The conceptualisation of educational assessment as symbolic media has significant analytic potential.

The conceptualisation of educational assessment as symbolic media with an intended interpretation and an actual interpretation, as well with a numerical and a linguistic component, has implications for system management. This study has found that scientific (Taylor, 1911/1998) and strategic (Jelinek, 1979; Mintzberg, 1994) forms of management are consistent with the outcome focus argued for by Masters (2005). This study has identified that an outcome focus, along with strategic and scientific forms of management, promotes numerical forms of system coordination in a manner where the numerical

component becomes uncoupled from its linguistic component. This study has suggested that the decoupling between linguistic and numerical components provides one explanation as to why Australia might be struggling on the PISA scales (Thomson et al., 2013; Thomson et al., 2017).

This study also suggests that a focus on the numerical component of educational assessment leads to a form of conservatism consistent with the disdain for progressive forces such as feminism expressed by Caldwell and Hayward (1998). This is because numerical coordination allows the values and norms expressed through educational assessment to become hidden, undefined, and removed from curriculum. This dynamic is captured by the way Friedman (1962/2002) uses the term “approved”. This approach leaves societal values unarticulated to reify existing social orders and reinforce what McIntosh (1986) terms white male privilege.

This study identifies that administrative management (Fayol, 1949/2013), with its regard for rules and processes (Weber, 1946/2009a), as having a greater potential to maintain connection between the numerical and linguistic components. Administrative forms of management can support institutionalisation of ethical discourse, deliberative democracy, and processes of justification and application identified by Habermas (1994, 1983/1996, 1992/1998). The coherent use of educational assessment is promoted when these processes are institutionalised as part of the integrative system identified by Parsons and Smelser (1956/2005).

Toulmin’s (1958/2003) approach to argument provides a fundamental underlying perspective on issues identified through this study around educational assessment design and use. Toulmin (1958/2003) describes two types of argument, one analytic and the other substantive. Analytical arguments are largely tautological, and this study has characterised the critique of Gorur and Wu (2015) on the use of ranks in policy-making as a critique of analytical arguments in educational assessment. That is, in the context of ranks and data, arguments are largely analytical and tautological as the justification for a claim is contained in the data. Substantial arguments are over what is true, what is right, and what is good. Substantial arguments require reference to broader evidence as well as referencing perceptions and feelings (Habermas, 1983/1996). It is at the level of substantial argument that recourse to further evidence is required, as well reference to

linguistic meaning-making. The approaches of Habermas (1983/1996) and Kane (2006), as well as of Mislavy, Steinberg, and Almond (2003), are all based on the notion of substantive argument as described by Toulmin (1958/2003).

How pragmatism relates to neoliberalism and postmodernism

This study has developed a pragmatic case and argument. Consistent with the psychological metaphor of the container and contained described by Bion (1963), this study has sought to be the container for holding neoliberalism and postmodernism in a dialogue on educational assessment. This study, from its pragmatic perspectives, argues that the other perspective can be incorporated through recourse to the better argument (Habermas, 1992/1998).

The central limitation of neoliberalism identified through this study relates to background understanding. Chapter 5 described how Habermas (1981/1985) considered Weber's (1915/1964) theory of action inadequate as it presupposes background understanding. This study considers neoliberalism limited in a similar way as it also presupposes a common understanding illustrated by Friedman's (1962/2002, p. 89) liberal use of the term "approved" towards education. Pragmatism is better able to frame what should be approved in education and educational assessment in pluralistic societies. The essence of Habermas' (1981/1985, 1981/1992, 1983/1996, 1992/1998) work addresses what is considered worthy of approval on behalf of all consistent with the pragmatic tradition that can be traced back to Plato's (1963) *Republic* (Popper, 1945/2002).

That neoliberalism explicitly eschews Plato (1963) in rejecting the collectivist tradition (Hayek, 1944/2007; Popper, 1945/2002) draws it into a performative contradiction (Habermas, 1983/1996; Watt, 1975). This contradiction emerges from the neoliberal argument that education should be universal and compulsory, as do Popper (1945/2002) and Friedman (1962/2002), but then refusing to engage in a conversation on what should constitute that education. Pragmatism can frame this conversation and there is an imperative on the neoliberal tradition to support and engage with it.

The postmodern tradition is more problematic from the perspective of this study. Postmodernism, with its focus on the subject, is also in performative contradiction (Habermas, 1983/1996; Watt, 1975) with respect to education. Ball (2016a, 2016b), for example, engages with neoliberalism in education but then refuses to engage with the

concept of neoliberalism. Ball (2013, p. 21) even declares that he is not so much interested in the field of education to proclaim that he is more interested in crafting an academic subject yet to come. Ball's (2013, 2016a, 2016b) position is hard to engage with as it is a critique that refuses to engage with concepts, refuses to declare its ambitions, and refuses to declare what it wants from society.

It is perhaps unfortunate that Habermas did not further elaborate developments in education and the knowledge economy (Murphy & Fleming, 2010) after the bifurcation between modernism and postmodernism (Habermas, 1981, 1997; Rorty, 1984, 1995). This study has shown that the concept of performativity developed by Lyotard (1979/1984) is useful in education. Given that it was developed and best articulated by Lyotard (1979/1984), it is also associated with the little narratives, an incredulity towards grand narratives, and the paralogy of Lyotard (1979/1984). This development has somewhat curtailed the responses to the empirical studies addressing performativity as described in Chapter 2. That Ball's (2013) conceptualisation of performativity underpins these studies has limited their impact as these critiques are not able to talk back to system perspectives and talk back to management processes inherent to education in contemporary societies.

The performative attitude (Habermas, 1992/1998; Weir, 1995) provides a counterpoint to Lyotard's (1979/1984) concept of performativity in a way that enhances its conceptual clarity. That is, where the study of Solomon and Lewin (2016), and the study of Hardy and Lewis (2016), identify contradictory logics arising out of performativity, these processes could also be couched in terms of the system-lifeworld model, where system imperatives colonise the lifeworld to restrict the performative attitude of students and teachers. A similar argument could be made about the work of Lambert et al. (2015) that identified a decline in arts subjects due to performativity. This study could also have been couched in terms of restricting the performative attitude in the three worlds to give a more classic treatment to the identified issues. For example, a diminishing focus on the arts could be cast into the context of the three critiques by Kant (1788/2004, 1790/2005, 1781/2016). The performative attitude provides a powerful concept in the critique of educational assessment and as a counterpoint to performativity.

The performative attitude has also been used by this study as a contrast to contemporary approaches to educational assessment on data collection. Where the OECD (2013c) and Masters (2005) address educational assessment in terms of generic student outcomes described by measures, the performative attitude orients educational assessment towards the objective, social and subjective worlds of the student. Furthermore, the performative attitude orients educational assessment as an inherent part of education regardless of data being generated and focuses educational assessment as an activity that is an investment in the student (Oakeshott, 1962/1991). Reorienting educational assessment towards the student challenges dominant contemporary approaches which generally consider educational assessment a process of extracting some form of data from the classroom for external stakeholders (Newton, 2007; OECD, 2013c). The performative attitude brings greater focus on educational assessment's role in the processes of socialisation, individuation, and the development of the self.

Implications for the theory of educational assessment validity

Through pragmatically integrating neoliberal, and postmodern perspectives with the more technically oriented field of educational assessment, significant implications emerge from this study. These implications arise from this study's exploration of human capital theory (G. S. Becker, 1964/1993) which links economic theory with educational theory using the respective media of money and educational assessment (Habermas, 1981/1992). The analysis in this study does not challenge the underpinning principles of human capital theory, instead it challenges the nature of the ratios made using measures of money and measures of educational achievement when each have distinct characteristics. Money is largely an abstract media and educational assessment a form of generalised communication that remains tied to linguistic content (Habermas, 1981/1992).

The implications for educational assessment emerge from the two conceptualisations of media. Money is an abstract steering media towards which users can take an empirically motivated, calculative, and purposive-rational attitude. When money is used in exchange there is no need for anyone to take responsibility for its meaning, and its exchange value can be indefinitely extended through autonomous and anonymous markets. In contrast, the media of educational assessment is a generalised form of communication that has a linguistic and a numerical component, where the linguistic component addresses the

economic, social and cultural spheres of society. The validity of educational assessment therefore involves limited horizons where validity may diminish as horizons broaden. Alternatively, an educational assessment may broaden its validity horizons to lose cultural specificity. Several economic implications emerge from mixing these two units in the application of human capital theory.

Two macroeconomic challenges

Two implications from this study's research emerge at the macroeconomic level. The first arises from Friedman's (1962/2008) assumption that wants are fixed. This study has found this assumption empirically and theoretically unsustainable (Bourdieu, 1979/1984; Jones, 1982/1995; Meyer, 2000; Peters, 1959/1973). Nevertheless, while unsustainable from a lifeworld perspective, this study also finds that the assumption might have some efficacy in economic planning and economic modelling. While this study has not explored the mathematical economic models as such, it notes that neoliberal and monetarist policy is associated with economic stability (OECD, 2017a). This observation is consistent with the framing of crises developed by Habermas (1975/2005), which distinguishes between economic crises and legitimation crises.

Nevertheless, this study finds that the current legitimation crises in educational assessment largely arise from the economic assumption that wants are fixed. This assumption promotes a realist stance towards the latent variable, and the use of numerical outcomes measures for coordination in the broader economy (Hanushek, 2013, 2016; Hanushek et al., 2013; Hanushek & Wößmann, 2010; Masters, 2005). Therefore, the first implication at the macroeconomic level from this study is that alternative economic models need to be developed which do not assume wants to be fixed and which are better able to address legitimation crises in educational assessment. This implication is unlikely to be a return to Keynesian (1953/1964) economics and is more likely to be an evolution of monetarist approaches that makes further recourse to postconventional reasoning (Friedman, 1953/2008, 1969/2008, 1962/2008; Habermas, 1983/1996; Kohlberg, 1971).

The second macroeconomic implication relates to the institutionalisation of educational assessment as symbolic media. The media of money is institutionally anchored through banks and governed by reserve banks in various jurisdictions. In the United States, for example, the *Federal Reserve* is responsible for influencing monetary and credit policy,

supervising institutions, maintaining financial stability, and facilitating the nation's payment system (Federal Reserve Board, 2005, p. 1). This level of centralised scrutiny and management is not evident in the design and use of educational assessment. Instead, this study observes markets in the creation of the symbolic media of educational assessment where these markets are not governed by openly declared standards.

Where Habermas (1981/1992, p. 180) argues symbolic media needs to have strategic action differentiated out, which institutions like the *Federal Reserve* (Federal Reserve Board, 2005) do for money, educational assessment has not yet reached this level of institutionalisation. Instead, the field of educational assessment is experiencing a proliferation of symbolic media (Edwards et al., 2012) in a fashion that somewhat corresponds to the proliferation of cryptocurrencies that emerge independent of the nation-state (Brenig et al., 2015; Gandal & Halaburda, 2014). That is, educational assessment measures have not yet stabilised, and there is an argument as to whether they can be stabilised in the same way as money given education's intrinsic link to culture and language. The OECD (2016b, 2017b, 2017e) could be considered as somewhat institutionalising and stabilising educational assessment measures, but this study observes that programs such as the PISA reify content and lack cultural specificity. Measures such as those generated by the PISA and NAPLAN also have difficulty responding to technological and social progress (Drucker, 1968/1992; Durkheim, 1933/2012; Levy & Murnane, 2004; OECD, 2010b). The challenge is therefore to develop conceptualisations of educational assessment that reduce the legitimisation tensions generated by programs such as the PISA and NAPLAN, while still meshing with measures of money in the application of human capital theory (G. S. Becker, 1964/1993; Blaug, 1976; Tan, 2014). The conceptualisation of symbolic media in educational assessment therefore provides an avenue for further research.

In summary, the second macroeconomic implication from this study pertains to the development of centralised educational measures that factor out strategic action of market actors, can incorporate cultural concerns, and reduce legitimisation tensions arising from use in economic planning. This is to reinforce the argument that educational assessment *ought* to be designed in recognition that education is in a dialectical relationship with the economy.

The challenge of moral development

A further challenge relates to moral development towards a postconventional level across societies (Kohlberg, 1971). Where the challenge at the macroeconomic level is to develop educational assessment symbolic media that is legitimate, this challenge is exacerbated because educational assessment is inherently a linguistically-based medium. This study has shown that it is often inappropriate to use educational assessment in a calculative and empirical sense. This view is reflected, for example, in the criticism of the ATAR in Australia by Finkel (2018). The ATAR, like all educational assessment, is a symbolic media that addresses the cognitive dimension of what students know and can do, how these relate to the normative context of society, and how these relate to the desires of the individual student. Where Newton (2012), for example, considers that claims arising from use of educational assessment data requires recourse to validity arguments, Finkel (2018) observes that universities and students are taking an increasingly purposive and empirical attitude towards the ATAR. The challenge is to create the ethical climate for reconnecting media, such as the ATAR, to the objective, social and subjective world of students through ordinary language. A challenge to ensure that the ATAR maintains a connection to a student's objectively measured capabilities, but also a connection to social norms and their personal disposition and desires.

Furthermore, the distinction Moss (2016) makes between intended interpretation and actual interpretation, and the call for a focus on how professionals used data, highlights the need for broad-based ethical development. The assessment validity literature, such as that explicated by Messick (1989) and Kane (2006), is framed in terms of "expert cultures ... [separated from] ... the communicative infrastructure of everyday life" (Habermas, 1981/1985, p. xxxii). What follows from Moss (2016), and consistent with the notion of discourses of application described by Habermas (1994), is that the capacity for postconventional reasoning, and for mounting substantive arguments about what is true, right and beautiful, needs to be diffused among those using educational assessment to make decisions. This includes policy-makers, judges, teachers, and the like (Moss, 2016; Shepard, 2016). The effective design and use of educational assessment require the development of ethical skills not only at the institutional level, but across society more broadly. This is to enable educational assessment to motivate rationally through trust,

rather than through purposive empirical ties (Habermas, 1981/1992, p. 183), at the level of Moss' (2016) actual interpretation.

Limitations of this study – the ideal speech situation

The approach to educational assessment validation emerging from this study is somewhat aspirational. This follows from the approach of Habermas (1983/1996) which is somewhat idealistic through notions such as the “ideal speech situation” (p. 88) and “ideal role taking” (p. 163). While this study has sought to identify and describe the underlying structures of educational assessment through immanent critique, this does not describe the actual use of educational assessment in broader society. Neither does this study seek to suggest a form of historicism where valid design and use of educational assessment is an inevitability of social progress. Nevertheless, this study does suggest that ongoing efforts towards the ideals expressed through this study would enhance the legitimacy of accountability regimes in education and the legitimacy of governments in general to create more stable societies.

Alternative scenarios suggest themselves with unpredictable consequences. Continued pursuit of neoliberalism in educational assessment would lead to a proliferation of educational assessment as symbolic media through markets. As explored by this study, market-based symbolic media can lead to more effective selection in university admissions (Edwards et al., 2012). However, following Kotler's (1972) axioms, markets only exist where there is an interest and the resources to participate. This is to suggest that market-based symbolic media is likely to lead to increased social balkanisation in society based on interest and available resources when compared to universal processes such as the ATAR. How this balkanisation might affect the stability of societies as a whole is not explored by this study.

The idealised approach toward validity by Habermas (1981/1985, 1983/1996) is distinct from Luhmann (1992, 1984/1995), for example, who considers symbolic media simply as non-random variation in social relations (Chernilo, 2002, p. 433). In this sense, communication between autonomous systems simply contributes to an autopoiesis between systems regardless of any claims to truth or validity. In this framing, students' maladaptive responses to NAPLAN, as identified by Howell (2017), can be considered as students seeking to maintain an internally coherent view of the world in response to

miscommunication. No imperative for an overarching value judgement emerges from the work of Luhmann (1992, 1984/1995), who might suggest these systems will self-correct over time, or simply cease to exist, without the need for the idealised deliberative processes advocated by Habermas (1992/1998).

The postmodern tradition, particularly that of Baudrillard (1981/1994), suggest the possibility that the media generated by educational assessment might become completely decoupled from reality. Further, the postmodern tradition such as that expressed by Ball (2013, 2016a, 2016b) advocates for a simple refusal to engage with these media, and a refusal to accept the norms they seek to represent. This approach evokes the behaviourist method of extinction (Skinner, 1938), to suggest that inappropriate forms of educational assessment will become extinct if simply ignored for long enough, regardless of deliberative processes (Habermas, 1992/1998).

Habermas (1992/1998), and this study more generally, has approached the topic through idealised institutionalised processes that promote a unity of purpose (Fayol, 1949/2013). In contrast, the postmodern approach of Ball (2016a, 2016b), and the autopoietic system approach of Luhmann (1992, 1984/1995), are both consistent with neoliberal approaches. Each avoid deliberative processes and their associated bureaucracies and rules, in a similar manner eschewed by the neoliberal tradition (Friedman, 1962/2002; Hayek, 1944/2007; Popper, 1945/2002; von Mises, 1944/2007). This study has not addressed the efficacy of self-managed and self-regulating educational assessment systems when compared to idealised ones based on principles of deliberative democracy (Habermas, 1992/1998).

Future research

Research imperatives arising from this study focus on methods and practices around constructs, mathematical models, test development and reporting. There are also research imperatives for system management and governance processes for institutionalising practices around these methods.

This study has identified the latent variable of educational assessment a social construct that is fluid and dependent on the context and purpose of an assessment (McCurry, 2017; Wiliam, 2010; Wu et al., 2017). This study has also identified that constructs address the three worlds and are subject to normative influences. This study has used Fensham (2016)

to illustrate how the construct of science can be influenced by considerations of gender. This concept can be extended to notions of sexuality (Foucault, 1984/1988, 1976/1990, 1984/1990), race (Gandhi, 1998; Said, 1978/1994; Spivak, 1985/2010) or any other category, particularly marginalised categories. Wiliam (2010), for example, argues that construct definitions interact with issues of equity in educational assessment. Research emerging from this study relate to what kind of processes would lead to fair and equitable construct definitions.

This study has shown a relationship between neoliberalism, the realist stance, and numerical methods for steering systems. Regardless of any philosophical intent underpinning these approaches, this study has identified that through these features neoliberalism reinforces a status quo, or what some call white male privilege (McIntosh, 1986). It also reinforces a status quo on matters of skills used in the economy. This study illustrated this relationship through the context of Victoria, where a pluralist educational climate emerged from the 1970s and into the 1980s (Barcan, 1996, 2003), which was subsequently curtailed during the 1990s under the influence of neoliberalism (Caldwell & Hayward, 1998; Fensham, 2016; Stacey, 2001). Research emerging from this study includes working towards identifying ways to create inclusive curriculum and assessment regimes that might accommodate the majority of citizens in society while not compromising economic prosperity.

On the matter of pluralism in the development of constructs, Habermas (1994, p. 105), following Gadamer (1976/2008, 1975/2013), alerts to the possibility of false alternatives, such as between assimilation and conversion. Instead, Habermas (1994) points to the potential for convergence between perspectives. It is in the definition of constructs that postconventional reasoning (Kohlberg, 1971) and perspective convergence becomes important. As argued by Wiliam (2010), it is at the level of construct that matters of equity and inclusion are best addressed, to suggest that further research in the composition of educational constructs might be fruitful for exploring issues of equity.

This study has also identified an emerging area of research around mathematical models. The ruler metaphor (Wright, 1997), and continuum model (Reckase, 2017), provide suitable mathematical models for some forms of psychological assessment and for providing system level statistics. Wu (2016), for example, suggests matters such as

differences between geolocations are able to be explored through the ruler metaphor. The ruler metaphor can also be imprecise and lacking in granularity for decision-making at the student level. The ruler metaphor is currently the dominant paradigm in Australian educational assessment through programs such as NAPLAN and the PISA (ACARA, 2015a; OECD, 2017e). This study has identified the constellation metaphor as potentially enhancing the granularity of evidence for decision-making at the classroom and school levels. A program of research potentially emerges around the use of the constellation metaphor, particularly as Reckase (2017) points to differences in test development for the two metaphors.

The importance of the constellation metaphor emerges from, what a recent Gonski et al. (2018) report describes as, the need to make more formative assessment available to teachers and schools. While this study identifies the term formative assessment unhelpful, the term is associated with notions such as normative validity, classroom-based assessment, and the lifeworld, as well as the performative attitude. It is a form of assessment that contrasts to summative assessment designed to fulfil system imperatives and which is associated with performativity. This study identifies the ruler metaphor as being more suitable for system level assessments, and the constellation metaphor more suitable for lifeworld or classroom-based assessments. This study therefore identifies a place for both metaphors in the suite of educational assessments used within systems. However, the relationship between the two would lead to the development of two distinct educational assessment programs that require a dialectic relationship between them. The constellation metaphor is therefore not just a technical solution, or a technical improvement, there are also broader political and administrative implications that need to be researched and worked through to make a formative assessment platform feasible, effective and legitimate.

This study has highlighted the importance of item writing and test development. These are important not just to the processes of educational assessment, but important for setting society's norms. Mislavy, Steinberg, and Almond (2003) associate educational assessment as communicating standards, expectations and values. Test development is also consistent with Habermas' (1983/1996) discourse ethics and principles of universalisation. That is, item writers develop content that can be accepted by all and where the norms encapsulated by an item are acceptable to all. Where legislators establish

norms through the medium of power and law, item writers establish norms through the linguistic medium of educational assessment. Research emerging from the growing importance of item writers in establishing valid norms includes matters such as the recruitment and selection of item writers, as well as their professional development. While item writers are important in setting standards and expectations, as observed by Kane (2006), item writers get a lot of latitude in this area. In this sense, development might also include item writers becoming more involved in public processes of justification and application (Habermas, 1994).

There are also emerging areas of research arising from item writers being implicated in the use of enhanced semiotic resources afforded by technology (Kress, 2003, 2010). As observed by Beavis (2010), new semiotic resources are likely to continue to affect the nature of communication around domains of knowledge used in society in a way that is likely to affect perceptions of heritage, culture and identity. It is with respect to new semiotic resources that item writers have a particular role in establishing and maintaining new norms and expectations on how society communicate through domains of knowledge. Related to these developments, is that item writers will need to increasingly work with game designers, learning designers, and assessment designers, as well as a range of technical areas such as video and animation production teams. The task of item writing is therefore likely to become more involved, suggesting an area of research around job descriptions and professional development.

Through developing the concept of educational assessment as symbolic media with intended and actual interpretations, this study has identified a research imperative around educational assessment reporting. The precepts for symbolic media developed by Parsons (1963a) suggests one starting point. Other factors to consider are those related to performativity, perlocutionary effects, and the creation of identity (Austin, 1962/1975; Butler, 1988, 1990/2007; Lyotard, 1979/1984). The maladaptive responses identified in the study by Howell (2017) presents one imperative for this type of work. Further work also emerges from the research agenda outlined by Moss (2016), which calls for greater support for professionals in making actual interpretations. Practices to support local interpretation include making test material publicly available. Other strategies include making publicly available supplementary material such as examiner reports, assessment frameworks, curriculum frameworks, and support for secondary analyses. This study

suggests that reporting, in terms of Austin's (1962/1975) illocutionary force, is a particularly important part of the educational assessment process and worthy of enhanced conceptualisation.

Finally, through identifying different paradigms towards validity for educational assessment and psychological assessment, this study has identified an emerging area of research into developing new standards for educational assessment. For example, the PISA Technical Standards (OECD, 2017f) explicitly orient educational assessment as a sociocultural practice influenced by political consideration. Further, the standards for the PISA address the notion of well-formed representative forums providing input into construct definitions and item development in a manner consistent with principles of deliberative democracy (Habermas, 1992/1998). Nevertheless, the standards used in the PISA which reports at the jurisdictional level are not appropriate for programs such as NAPLAN which reports at the student level. There is therefore an emerging area of research into standards that explicitly address educational assessment as a sociocultural activity that is oriented towards the economy and polity.

A broad agenda for future research is suggested by this study with several branches proceeding proceeds in different directions, each direction encountering a variety of disciplines. This study has sought to synthesise in the tradition of Hegel (1807/1977, 1816/1998), and sits somewhere between philosophy and practice, and between theory and practice (Habermas, 1973/2007). In doing so the specificity of the research agenda suggested by this study is somewhat lost. Two broad areas of dissemination might therefore emerge from these efforts, a book focusing on dissemination to a general public about a general theory, and targeted research articles that detail specific elements in reference to the general theory to foster specific research. Such a strategy would simultaneously address the integration of progress narratives into a whole, and its differentiation into specialised areas of work.

The general theory emerging from this study potentially affects many stakeholders, including: intergovernmental agencies such as the OECD; governments and bureaucracies; schools and school systems; public and private providers; and technology providers. As this study does not seek to prescribe how educational assessment is

designed and used, to instead leave this to deliberative processes, determining how stakeholders might be affected becomes part of these processes.

Conclusion

While this study has identified extensive critiques and issues in educational assessment, there is also a sense that it is a field that will continue to emerge in importance. Measures of human capital, as well as measures of social capital, will continue to be important for governments seeking to maximise citizen engagement and participation in economies and societies. Educational assessment will also remain important in terms of allowing students, as citizens, to individuate along cognitive, moral, and aesthetic dimensions through the performative attitude. These aspects are likely to become increasingly important for advanced economies where knowledge is paramount and where measures of knowledge are essential for planning and coordination.

System imperatives for planning and coordination will continue to generate demand for high-stakes large-scale educational assessment regimes. These system driven regimes can become increasingly isolated from educational assessment practices addressing the emancipatory interest of students through, for example, care-referenced and student referenced concerns (Cowie & Bell, 1999). Flórez Petour (2015) characterises this emancipatory interest as “embedded in a long-standing ideological struggle, where high-stakes assessment systems ... keep innovative repertoires in the margins of the system” (p. 20). It is towards this ideological struggle that this study has spoken.

References

- ACARA. (2014). *National Assessment Program – Literacy and Numeracy: Technical report 2013*. Sydney: Australian Curriculum, Assessment and Reporting Authority. Retrieved from www.acara.edu.au
- ACARA. (2015a). *NAPLAN 2015 technical report*. Sydney: Australian Curriculum, Assessment and Reporting Authority. Retrieved from www.acara.edu.au
- ACARA. (2015b). *National Assessment Program - Literacy and Numeracy: Achievement in reading, persuasive writing, language conventions and numeracy: National report for 2015*. Sydney: Australian Curriculum, Assessment and Reporting Authority. Retrieved from www.acara.edu.au
- ACARA. (2018). Release of NAPLAN 2018 summary information [Press release]. Retrieved from <http://www.acara.edu.au/news-and-media/media-releases>
- Achrol, R. S., & Kotler, P. (2011). Frontiers of the marketing paradigm in the third millennium. *Journal of the Academy of Marketing Science*, 40(1), 35-52. <https://doi.org/10.1007/s11747-011-0255-4>
- Adams, R. J. (2002). Scaling PISA cognitive data. In R. J. Adams & M. Wu (Eds.), *PISA 2000 technical report* (pp. 99-108). Paris: OECD Publishing. <https://doi.org/10.1787/9789264199521-en>
- Adams, R. J., Wilson, M., & Wang, W.-c. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21(1), 1-23. <https://doi.org/10.1177/0146621697211001>
- Adams, R. J., & Wu, M. (2007). The mixed-coefficients multinomial logit model: A generalized form of the Rasch model. In M. v. Davier & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models: Extensions and applications* (pp. 57-75). New York: Springer.
- Adams, R. J., & Wu, M. (Eds.). (2002). *PISA 2000 technical report*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264199521-en>
- AERA/APA/NCME. (2014). *Standards for educational and psychological testing*. Washington DC: American Educational Research Association, American Psychological Association, & National Council on Measurement in Education.
- Alexander, C., Ishikawa, S., Silverstein, M., Jacobson, M., Fiksdahl-King, I., & Angel, S. (1977). *A pattern language: Towns, buildings, construction*. New York: Oxford University Press.

- Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31(2), 211-236. <https://doi.org/10.1257/jep.31.2.211>
- Almond, R. G., Mislevy, R. J., Steinberg, L. S., Yan, D., & Williamson, D. M. (2015). *Bayesian networks in educational assessment*. New York: Springer.
- Althaus, C., Bridgman, P., & Davis, G. (2013). *The Australian policy handbook* (5th ed.). Sydney: Allen & Unwin. (Original work published 1998)
- American Educational Research Association. (2015). AERA Statement on use of Value-Added Models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher*, 44(8), 448-452. <https://doi.org/10.3102/0013189x15618385>
- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, 84(3), 261-271. <https://doi.org/10.1037/0022-0663.84.3.261>
- Ananiadou, K., & Claro, M. (2009). *21st century skills and competences for new millennium learners in OECD countries* (OECD Education Working Papers No. 41). Paris: OECD Publishing. <https://doi.org/10.1787/218525261154>
- Anderson, J. (2011). Autonomy, agency and self. In B. Fultner (Ed.), *Jürgen Habermas: Key concepts* (pp. 91-112). Durham: Acumen.
- Anderton, R. S., Evans, T., & Chivers, P. T. (2016). Predicting academic success of health science students for first year anatomy and physiology. *International Journal of Higher Education*, 5(1). <https://doi.org/10.5430/ijhe.v5n1p250>
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43(4), 561-573.
- Andrich, D. (1988). *Rasch models for measurement*. Newbury Park CA: SAGE Publications.
- Antonio, R. J. (1981). Immanent critique as the core of critical theory: Its origins and developments in Hegel, Marx and contemporary thought. *The British Journal of Sociology*, 32(3), 330-345. <https://doi.org/10.2307/589281>
- AQFC. (2013). *Australian Qualifications Framework: Second edition January 2013*. South Australia: Australian Qualifications Framework Council. Retrieved from www.aqf.edu.au

- ARG. (1999). *Assessment for learning: Beyond the Black Box*. London: The Assessment Reform Group. Retrieved from http://www.nuffieldfoundation.org/sites/default/files/files/beyond_blackbox.pdf
- Arnstein, S. R. (1969). A ladder of citizen participation. *Journal of the American Institute of Planners*, 35(4), 216-224. <https://doi.org/10.1080/01944366908977225>
- Aune, J. A. (2007). "Only Connect": Between Morality and Ethics in Habermas' Communication Theory. *Communication Theory*, 17(4), 340-347. <https://doi.org/10.1111/j.1468-2885.2007.00300.x>
- Austin, J. L. (1975). *How to do things with words* (J. O. Urmson & M. Sbisá Eds. 2nd ed.). Cambridge MA: Harvard University Press. (Original work published 1962)
- Bacon, M. (2012). *Pragmatism: An introduction*. Cambridge UK: Polity.
- Bagnato, S. J., Goins, D. D., Pretti-Frontczak, K., & Neisworth, J. T. (2014). Authentic assessment as "best practice" for early childhood intervention: National consumer social validity research. *Topics in Early Childhood Special Education*, 34(2), 116-127. <https://doi.org/10.1177/0271121414523652>
- Baird, J.-A., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment and learning: fields apart? *Assessment in Education: Principles, Policy & Practice*, 24(3), 317-350. <https://doi.org/10.1080/0969594X.2017.1319337>
- Baird, J.-A., Cresswell, M., & Newton, P. E. (2000). Would the real gold standard please step forward? *Research Papers in Education*, 15(2), 213-229. <https://doi.org/10.1080/026715200402506>
- Baird, J.-A., & Lee-Kelley, L. (2009). The dearth of managerialism in implementation of national examinations policy. *Journal of Education Policy*, 24(1), 55-81. <https://doi.org/10.1080/02680930802382938>
- Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), *Learning analytics: From research to practice* (pp. 61-75). New York: Springer.
- Ball, S. J. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy*, 18(2), 215-228. <https://doi.org/10.1080/0268093022000043065>
- Ball, S. J. (2013). *Foucault, power, and education*. New York: Routledge.

- Ball, S. J. (2015). Education, governance and the tyranny of numbers. *Journal of Education Policy*, 30(3), 299-301.
<https://doi.org/10.1080/02680939.2015.1013271>
- Ball, S. J. (2016a). Neoliberal education? Confronting the slouching beast. *Policy Futures in Education*, 14(8), 1046-1059.
<https://doi.org/10.1177/1478210316664259>
- Ball, S. J. (2016b). Subjectivity as a site of struggle: refusing neoliberalism? *British Journal of Sociology of Education*, 37(8), 1129-1146.
<https://doi.org/10.1080/01425692.2015.1044072>
- Barcan, A. (1996). Attempts to reform Australian state schools, 1979-1996. *Education Research and Perspectives*, 23(1), 1-45.
- Barcan, A. (2003). The struggle for curriculum reform in Australia 1987-1993. *Education Research and Perspectives*, 30(2), 108-139.
- Bardach, E. (2012). *A practical guide for policy analysis: The eightfold path to more effective problem solving* (4th ed.). Los Angeles: Sage.
- Baron-Cohen, S., Murphy, L., Chakrabarti, B., Craig, I., Mallya, U., Lakatošová, S., . . . Allison, C. (2014). A genome wide association study of mathematical ability reveals an association at chromosome 3q29, a locus associated with autism and learning difficulties: a preliminary study. *PloS one*, 9(5), e96374.
<https://doi.org/10.1371/journal.pone.0096374>
- Barthes, R. (1977). *Image music text* (S. Heath, Trans.). London: Fontana Press.
- Barthes, R. (1993). *Mythologies* (A. Lavers, Trans.). London: Vintage. (Original work published 1957)
- Baudrillard, J. (1994). *Simulacra and simulation* (S. F. Glaser, Trans.). Ann Arbor: University of Michigan Press. (Original work published 1981)
- Beare, H. (1983). The structural reform movement in Australian education during the 1980s and its effect on schools. *Journal of Educational Administration*, 21(2), 149-168. <https://doi.org/10.1108/eb009875>
- Beavis, C. A. (2010). English in the digital age: Making English digital. *English in Australia*, 45(2), 21-30. Available from
<http://www98.griffith.edu.au/dspace/handle/10072/37149>

- Becker, G. S. (1993). *Human capital: A theoretical and empirical analysis, with special reference to education* (3rd ed.). Chicago IL: University of Chicago Press. (Original work published 1964)
- Becker, K. A., & Bergstrom, B. A. (2013). Test administration models. *Practical Assessment, Research & Evaluation, 18*(14), 7-7. Available from <http://pareonline.net/pdf/v18n14.pdf>
- Behrens, J. T., & DiCerbo, K. E. (2013). Technological implications for assessment ecosystems (pp. 101-122). Princeton, NJ: The Gordon Commission on the Future of Assessment.
- Behrens, J. T., & DiCerbo, K. E. (2014). Harnessing the currents of the digital ocean. In J. A. Larusson & B. White (Eds.), *Learning analytics: From research to practice* (pp. 39-60). New York: Springer.
- Beiser, F. C. (2005). *Hegel*. New York: Routledge
- Berenson, F. (1982). Hegel on others and the self. *Philosophy, 57*(219), 77-90. Available from <http://www.jstor.org/stable/4619540>
- Berson, I. R., & Berson, M. J. (2006). Children and their digital dossiers: Lessons in privacy rights in the digital age. *International Journal of Social Education, 21*(1), 135-147.
- Biesta, G. (2009). Good education in an age of measurement: On the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability, 21*(1), 33-46. <https://doi.org/10.1007/s11092-008-9064-9>
- Biesta, G. (2010). *Good education in an age of measurement*. London: Routledge.
- Biesta, G. (2015a). Education, measurement and the professions: Reclaiming a space for democratic professionalism in education. *Educational Philosophy and Theory, 49*(4), 315-330. <https://doi.org/10.1080/00131857.2015.1048665>
- Biesta, G. (2015b). Resisting the seduction of the global education measurement industry: notes on the social psychology of PISA. *Ethics and Education, 10*(3), 348-360. <https://doi.org/10.1080/17449642.2015.1106030>
- Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 17-66). Dordrecht: Springer.

- Bion, W. R. (1963). *Elements of psycho-analysis*. London: William Heinemann Medical Books.
- Bion, W. R. (1996). *Experiences in groups: and other papers*. London: Routledge. (Original work published 1961)
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 92(1), 81-90.
<https://doi.org/10.1177/003172171009200119>
- Black, P., & Wiliam, D. (2003). 'In praise of educational research': Formative assessment. *British Educational Research Journal*, 29(5), 623-637.
<https://doi.org/10.1080/0141192032000133721>
- Blaug, M. (1976). The empirical status of human capital theory: A slightly jaundiced survey. *Journal of Economic Literature*, 14(3), 827-855.
- Blum, W. (1993). Mathematical modelling in mathematics education and instruction. In I. Huntley, T. Breiteig, & G. Kaiser-Messmer (Eds.), *Teaching and learning mathematics in context* (pp. 3-14). New York: Ellis Horwood.
- Blundell-Wignall, A. (2014). *The Bitcoin question: Currency versus trust-less transfer technology* (OECD Working Papers on Finance, Insurance and Private Pensions No. 37). OECD Publishing. <https://doi.org/10.1787/5jz2pwjd9t20-en>
- Blyth, K. (2014). Selection methods for undergraduate admissions in Australia. Does the Australian predominate entry scheme the Australian Tertiary Admissions Rank (ATAR) have a future? *Journal of Higher Education Policy and Management*, 36(3), 268-278. <https://doi.org/10.1080/01587919.2014.899049>
- Bonal, X., & Tarabini, A. (2013). The role of PISA in shaping hegemonic educational discourses, policies and practices: The case of Spain. *Research in Comparative and International Education*, 8(3), 335-341.
<https://doi.org/10.2304/rcie.2013.8.3.335>
- Boomsma, A., van Duijn, M., & Snijders, T. (Eds.). (2001). *Essays on item response theory* (Vol. 157). New York: Springer-Verlag.
- Borsboom, D. (2008). Latent variable theory. *Measurement: Interdisciplinary Research and Perspectives*, 6(1-2), 25-53. <https://doi.org/10.1080/15366360802035497>
- Borsboom, D., & Markus, K. A. (2013). Truth and evidence in validity theory. *Journal of Educational Measurement*, 50(1), 110-114.
<https://doi.org/10.1111/jedm.12006>

- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*(2), 203-219. <https://doi.org/10.1037/0033-295X.110.2.203>
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071. <https://doi.org/10.1037/0033-295X.111.4.1061>
- Borsboom, D., & Wijsen, L. D. (2016). Frankenstein’s validity monster: the value of keeping politics and science separated. *Assessment in Education: Principles, Policy & Practice*, *23*(2), 281-283. <https://doi.org/10.1080/0969594x.2016.1141750>
- Borsboom, D., & Wijsen, L. D. (2017). Psychology’s atomic bomb. *Assessment in Education: Principles, Policy & Practice*, *24*(3), 440-446. <https://doi.org/10.1080/0969594x.2017.1333084>
- Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste*. Cambridge MA: Harvard University Press. (Original work published 1979)
- Bourdieu, P. (1986). The forms of capital.
- Bower, J. L., & Christensen, C. M. (1995). Disruptive technologies: catching the wave. *Harvard Business Review*, *73*(1), 43-53. Available from <https://hbr.org/1995/01/disruptive-technologies-catching-the-wave>
- Bowie, A. (2003). *Introduction to German philosophy: From Kant to Habermas*. Cambridge UK: Polity.
- Bowman, S., McKinstry, C., & McGorry, P. (2017). Youth mental ill health and secondary school completion in Australia: time to act. *Early Intervention in Psychiatry*, *11*(4), 277-289. <https://doi.org/10.1111/eip.12357>
- Braaten, J. (1991). *Habermas's critical theory of society*. Albany: State University of New York Press.
- Braaten, J. (1995). From communicative rationality to communicative thinking. In J. Meehan (Ed.), *Feminists read Habermas: Gendering the subject of discourse* (pp. 139-167). New York: Routledge.
- Braun, H. (2003). Assessment and technology. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards* (pp. 267-288). Dordrecht: Springer. https://doi.org/10.1007/0-306-48125-1_11

- Breakspear, S. (2012). *The policy impact of PISA: An exploration of the normative effects of international benchmarking in school system performance*. (OECD Education Working Papers No. 71). Paris: OECD Publishing.
<https://doi.org/10.1787/5k9fdfqffr28-en>
- Brenig, C., Accorsi, R., & Müller, G. (2015). *Economic analysis of cryptocurrency backed money laundering* (ECIS 2015 Completed Research Papers No. 20).
<https://doi.org/10.18151/7217279>
- Brennan, R. L. (2006a). Perspectives on the evolution and future of educational measurement. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 1-16). Westport, CT: American Council on Education and Praeger Publishers.
- Brennan, R. L. (Ed.) (2006b). *Educational measurement* (4th ed.). Westport, CT: American Council on Education and Praeger Publishers.
- Brown, G. T., & Harris, L. R. (Eds.). (2016). *Handbook of human and social conditions in assessment*. New York: Routledge.
- Bryant, D. A., & Walker, A. (2016). A review of the linkage between student participation in the International Baccalaureate Continuum and student learning attributes. *Journal of Research in International Education*, 15(2), 87-105.
<https://doi.org/10.1177/1475240916658743>
- Buchanan, I. (2010). *A Dictionary of Critical Theory* (first ed.). Oxford: Oxford University Press.
- Buchanan, R., Southgate, E., Scevak, J., & Smith, S. P. (2018). Expert insights into education for positive digital footprint development. *Scan*, 37(2). Available from <https://education.nsw.gov.au/teaching-and-learning/professional-learning/scan/latest-issue/expert-insights-into-education-for-positive-digital-footprint-development>
- Bulkley, K. E., & Burch, P. (2011). The changing nature of private engagement in public education: For-profit and nonprofit organizations and educational reform. *Peabody Journal of Education*, 86(3), 236-251.
<https://doi.org/10.1080/0161956X.2011.578963>
- Burch, P. (2009). *Hidden Markets: The New Education Privatization*. New York: Routledge.
- Burch, P., & Good, A. G. (2014). *Equal Scrutiny: Privatization and Accountability in Digital Education*. Cambridge MA: Harvard Education Press.

- Butler, J. (1988). Performative acts and gender constitution: An essay in phenomenology and feminist theory. *Theatre Journal*, 40(4), 519-531. <https://doi.org/10.2307/3207893>
- Butler, J. (2007). *Gender trouble: Feminism and the subversion of identity*. New York: Routledge. (Original work published 1990)
- Caldwell, B. J., & Hayward, D. K. (1998). *The future of schools: Lessons from the reform of public education*. London: Falmer Press.
- Caldwell, B. J., & Spinks, J. M. (1988). *The self-managing school*. London: The Falmer Press.
- Callahan, R. E. (2007). *Education and the cult of efficiency: A study of the social forces that have shaped the administration of the public schools*. Chicago University of Chicago Press. (Original work published 1962)
- Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 221-256). Westport, CT: American Council on Education and Praeger Publishers.
- Capon, N. (1996). [Review of the book *The Rise and Fall of Strategic Planning*, by H. Mintzberg]. *The Academy of Management Review*, 21(1), 298-301. Available from <https://www.jstor.org/stable/258641>
- Carless, S. A. (2007). Graduate recruitment and selection in Australia. *International Journal of Selection and Assessment*, 15(2), 153–166. <https://doi.org/10.1111/j.1468-2389.2007.00377.x>
- Carnap, R. (1962). *Logical foundations of probability*. Chicago: The University of Chicago Press.
- Carnap, R. (2005). *The logical structure of the world: and, pseudoproblems in philosophy* (R. A. George, Trans.). Chicago: Open Court. (Original work published 1928)
- Chernilo, D. (2002). The theorization of social co-ordinations in differentiated societies: the theory of generalized symbolic media in Parsons, Luhmann and Habermas. *The British Journal of Sociology*, 53(3), 431-449. <https://doi.org/10.1080/0007131022000000581>
- Child Wellbeing and Safety Act 2005* (Vic) (Austl.). Available from http://www.austlii.edu.au/au/legis/vic/consol_act/cwasa2005218/
- Chomsky, N. (1969). *Aspects of the theory of syntax*. Cambridge MA: MIT press.

- Chomsky, N. (2015). *Syntactic structures*. Mansfield Centre CT: Martino Publishing. (Original work published 1957)
- Chung, J. H. (2010). Finland, PISA, and the implications of international achievement studies on education policy. In A. W. Wiseman (Ed.), *The impact of international achievement studies on national education policymaking* (pp. 267-294). Bingley: Emerald Group Publishing. [https://doi.org/10.1108/S1479-3679\(2010\)0000013013](https://doi.org/10.1108/S1479-3679(2010)0000013013)
- Civil Rights Act of 1964* (USA.). Available from http://crdl.usg.edu/events/civil_rights_act_1964/
- Cizek, G. J. (2012a). Defining and distinguishing validity: interpretations of score meaning and justifications of test use. *Psychological Methods*, 17(1), 31-43. <https://doi.org/10.1037/a0026975>
- Cizek, G. J. (2016). Progress on validity: The glass half full, the work half done. *Assessment in Education: Principles, Policy & Practice*, 23(2), 304-308. <https://doi.org/10.1080/0969594x.2016.1156642>
- Cizek, G. J. (Ed.) (2012b). *Setting performance standards: Foundations, methods, and innovations* (2nd ed.). New York: Routledge.
- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks: SAGE Publications.
- Cohen, A. S., & Wollack, J. A. (2006). Test administration, security, scoring and reporting. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 355-386). Westport, CT: American Council on Education and Praeger Publishers.
- Cole, N. S., & Moss, P. A. (1989). Bias in test use. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 201-219). Washington DC: American Council on Education.
- Coleman, J. S., & Hoffer, T. (1987). *Public and private high schools: The impact of communities*. New York Basic Books.
- Collis, B. A., & Rosenblood, L. K. (1985). The problem of inflated significance when testing individual correlations from a correlation matrix. *Journal for Research in Mathematics Education*, 16(1), 52-55. Available from <http://www.jstor.org/stable/748973>
- Cooke, M. (1997). *Language and reason: A study of Habermas's pragmatics*. Cambridge MA: MIT Press.

- Costa, P. T., & McCrae, R. R. (1993). Ego development and trait models of personality. *Psychological Inquiry*, 4(1), 20-23. Available from <http://www.jstor.org/stable/1449588>
- Cowie, B., & Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education: Principles, Policy & Practice*, 6(1), 101-116. <https://doi.org/10.1080/09695949993026>
- Crenshaw, K. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review*, 43(6), 1241-1299. Available from <http://www.jstor.org/stable/1229039>
- Crocker, L. (1997a). The great validity debate. *Educational Measurement: Issues and Practice*, 16(2), 4. <https://doi.org/10.1111/j.1745-3992.1997.tb00584.x>
- Crocker, L. (Ed.) Validity [Special issue]. *Educational Measurement: Issues and practices*, 16(2).
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443-507). Washington DC: American Council on Education.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, 52(4), 281-302. <https://doi.org/10.1037/h0040957>
- Cronin, C. (1994). Translator's introduction. In J. Habermas, *Justification and application: Remarks on discourse ethics*. Cambridge MA: MIT Press.
- Csikszentmihalyi, M. (2008). *Flow: The psychology of optimal experience*. New York: HarperPerennial. (Original work published 1990)
- Cuypers, S. E., & Martin, C. (2013). *RS Peters* (Vol. 18). London: Bloomsbury.
- Cuypers, S. E., & Martin, C. (Eds.). (2011). *Reading R. S. Peters today: Analysis, ethics, and the aims of education*. West Sussex: John Wiley & Sons.
- Daley, J., Duckett, S., Goss, P., Minifie, J., Norton, A., Terrill, M., & Wood, T. (2016). *Orange book 2016: Priorities for the next commonwealth government*: Grattan Institute.
- Davey, T., Pitoniak, M. J., & Slater, S. C. (2016). Designing computer adaptive tests. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds.), *The handbook of test development* (2 ed., pp. 467-484). New York: Routledge.

- Davies, A., & Goldsmith, S. M. (1988). Controlling for experimentwise errors: The Bonferroni procedure. *Reading Research Quarterly*, 23(3), 379-380. Available from <http://www.jstor.org/stable/748048>
- Davis, L. L., Strain-Seymour, E., & Gay, H. (2013). *Testing on tablets: Part II of a series of usability studies on the use of tablets for K-12 assessment programs*: Pearson. Retrieved from <https://www.pearson.com/corporate/efficacy-and-research/schools-education-research/research-reports/assessment/testing-on-tablets.html>
- Dawes, S. S. (2010). Stewardship and usefulness: Policy principles for information-based transparency. *Government Information Quarterly*, 27(4), 377-383. <https://doi.org/10.1016/j.giq.2010.07.001>
- Dawson, P. (2016). Five ways to hack and cheat with bring-your-own-device electronic examinations. *British Journal of Educational Technology*, 47(4), 592-600. <https://doi.org/10.1111/bjet.12246>
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: The Guilford Press.
- de Berg, H., & Large, D. (2012). *Modern German thought from Kant to Habermas: an annotated German-language reader*. New York: Camden House.
- de Certeau, M. (1984). *The practice of everyday life* (S. Rendall, Trans.). Berkeley: University of California Press.
- de la Torre, J., & Minchen, N. (2014). Cognitively diagnostic assessments and the cognitive diagnosis model framework. *Psicología Educativa*, 20(2), 89-97. <https://doi.org/10.1016/j.pse.2014.11.001>
- DEECD. (2011). *Maternal and child health service guidelines*. Victoria: Department of Education and Early Childhood Development. Retrieved from http://www.education.vic.gov.au/Documents/childhood/professionals/health/mc_hsguidelines.pdf
- Delić, H., & Bećirović, S. (2016). Socratic Method as an Approach to Teaching. *European Researcher*, 111(10). <https://doi.org/10.13187/er.2016.111.511>
- Department of Education and Training Victoria. (2017). *Working with integrity: The Department of Education and Training's second report to the Independent Broad-based Anti-Corruption Commission*. Melbourne: Department of Education and Training, Victoria.

- Department of Education and Training Victoria. (n.d). Safe Schools. Retrieved 11 October 2017 from <http://www.education.vic.gov.au/about/programs/health/Pages/safe-schools-coalition.aspx?Redirect=1#link31>
- Derrida, J. (1981). *Positions* (A. Bass, Trans.). Chicago: University of Chicago Press. (Original work published 1972)
- Derrida, J. (2007). Structure, sign, and play in the discourse of the human sciences. In R. Macksey & E. Donato (Eds.), *The structuralist controversy: The languages of criticism & the sciences of man* (40th Anniversary ed., pp. 247-272). Baltimore: Johns Hopkins University Press. (Original work published 1966)
- Derrida, J., & Caputo, J. D. (1997). *Deconstruction in a nutshell: A conversation with Jacques Derrida*. New York: Fordham University Press.
- Dewey, J. (2011). *Democracy and education*. Milton Keynes, UK: Simon & Brown. (Original work published 1916)
- Didau, D. (2016). *What if everything you knew about education was wrong?* Carmarthen: Crown House Publishing.
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, *41*(1), 417-440. <https://doi.org/10.1146/annurev.ps.41.020190.002221>
- Diquinzio, P. (1993). Exclusion and essentialism in feminist theory: The problem of mothering. *Hypatia*, *8*(3), 1-120. Available from <https://www.jstor.org/stable/3810402>
- Donaldson, T., & Preston, L. E. (1995). The stakeholder theory of the corporation: Concepts, evidence, and implications. *Academy of Management Review*, *20*(1), 65-91. <https://doi.org/10.5465/AMR.1995.9503271992>
- Dorans, N. J., Moses, T. P., & Eignor, D. R. (2011). Equating test scores: Toward best practices. In A. von Davier (Ed.), *Statistical models for test equating, scaling, and linking* (pp. 21-42). New York: Springer.
- Downing, S. M. (2006). Twelve steps for effective test development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 3-26). Mahwah NJ: Lawrence Erlbaum Associates.
- Drasgow, F., Luecht, R. M., & Bennett, R. E. (2006). Technology and testing. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 471-515). Westport, CT: American Council on Education and Praeger Publishers.

- Drucker, P. F. (1992). *The age of discontinuity: Guidelines to our changing society*. New Brunswick NJ: Transaction Publishers. (Original work published 1968)
- Drucker, P. F. (1994). *Post-capitalist society*. New York: HarperCollins.
- Dunn, O. J. (1959). Estimation of the Medians for Dependent Variables. *The Annals of Mathematical Statistics*, 30(1), 192-197. Available from <http://www.jstor.org/stable/2237135>
- Dunn, O. J. (1961). Multiple Comparisons Among Means. *Journal of the American Statistical Association*, 56(293), 52-64. Available from <http://www.jstor.org/stable/2282330>
- Durkheim, É. (2001). *The Elementary Forms of the Religious Life*. Oxford: Oxford Press.
- Durkheim, É. (2012). *The division of labor in society* (G. Simpson, Trans.). Mansfield Centre CT: Martino Publishing. (Original work published 1933)
- Eacott, S. (2017). School leadership and the cult of the guru: the neo-Taylorism of Hattie. *School Leadership & Management*, 1-14. <https://doi.org/10.1080/13632434.2017.1327428>
- Eagleton, T. (2008). *Literary theory: An introduction* (Anniversary ed.). Minneapolis MN: University of Minnesota Press. (Original work published 1983)
- Education and Training Reform Act 2006* (Vic) (Austl.). Available from http://www.austlii.edu.au/au/legis/vic/consol_act/etra2006273/
- Edwards, D., Coates, H., & Friedman, T. (2012). A survey of international practice in university admissions testing. In V. Massaro (Ed.), *Higher Education Management and Policy: Journal of the Programme on Institutional Management in Higher Education* (pp. 87-104). Paris: OECD Publishing. <https://doi.org/10.1787/hemp-v24-1-en>
- Egan, K. (1998). *The educated mind: How cognitive tools shape our understanding*. Chicago: University of Chicago Press.
- Ellery, S. (2017). Western Australia to delay NAPLAN online to 2018 [Press release]. Retrieved from <https://www.mediastatements.wa.gov.au/Pages/McGowan/2017/04/Western-Australia-to-delay-NAPLAN-Online-to-2018.aspx>

- Elliott, J. E. (1980). Marx and Schumpeter on Capitalism's Creative Destruction: A Comparative Restatement*. *The Quarterly Journal of Economics*, 95(1), 45-68. <https://doi.org/10.2307/1885348>
- Elliott, S. L., & Epstein, J. (2005). Selecting the future doctors: the role of graduate medical programmes. *Internal Medicine Journal*, 35, 174-177.
- Elliott, S. N., & Kettler, R. J. (2016). Item and test design considerations for students with special needs. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds.), *The handbook of test development* (2 ed., pp. 374-391). New York: Routledge.
- Ellis, V., & McNicholl, J. (2015). *Transforming Teacher Education: Reconfiguring the Academic Work*. London: Bloomsbury Publishing.
- Elmore, R. F. (1996). Getting to scale with good educational practice. *Harvard Educational Review*, 66(1), 1-27. <https://doi.org/10.17763/haer.66.1.g73266758j348t33>
- Elmore, R. F. (2016). "Getting to scale..." it seemed like a good idea at the time. *Journal of Educational Change*, 17(4), 529-537. <https://doi.org/10.1007/s10833-016-9290-8>
- Elwood, J. (2013). Educational assessment policy and practice: A matter of ethics. *Assessment in Education: Principles, Policy & Practice*, 20(2), 205-220. <https://doi.org/10.1080/0969594x.2013.765384>
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah NJ: Lawrence Erlbaum.
- Engelmann, S., Becker, W. C., Carnine, D., & Gersten, R. (1988). The direct instruction follow through model: Design and outcomes. *Education and Treatment of Children*, 11(4), 303-317. Available from <http://www.jstor.org/stable/42899079>
- Eskerod, P., & Huemann, M. (2013). Sustainable development and project stakeholder management: what standards say. *International Journal of Managing Projects in Business*, 6(1), 36-50. <https://doi.org/10.1108/17538371311291017>
- Fabo, B., & Drahokoupil, J. (2016). *The platform economy and the disruption of the employment relationship* (ETUI Policy Brief No. 5). Brussels: The European Union.
- Fairbairn, W. R. D. (1994). *Psychoanalytic studies of the personality*. London: Routledge. (Original work published 1952)

- Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). *Teaching Adolescents to Become Learners: The Role of Noncognitive Factors in Shaping School Performance--A Critical Literature Review*. Chicago: The University of Chicago Consortium on Chicago School Research.
- Fayol, H. (2013). *General and industrial management* (C. Storrs, Trans.). Mansfield Centre CT: Martino Publishing. (Original work published 1949)
- Federal Reserve Board. (2005). *The federal reserve system: Purposes and functions*. Washington DC: Board of Governors of the Federal Reserve System. Retrieved from www.federalreserve.gov
- Feng, M., & Heffernan, N. T. (2006). Informing teachers live about student learning: Reporting in the assistment system. *Technology Instruction Cognition and Learning*, 3(1/2), 115-128. Available from <http://www.oldcitypublishing.com>
- Fensham, P. J. (2016). The future curriculum for school science: What can be learnt from the past? *Research in Science Education*, 46(2), 165–185. <https://doi.org/10.1007/s11165-015-9511-9>
- Fernald, D. (2007). *Psychology: six perspectives*. Los Angeles: Sage Publications.
- Finkel, A. (2018). *Optimising STEM industry-school partnerships: Inspiring Australia's next generation*. Carlton South: Education Council Secretariat.
- Finn, B. (1991). *Young people's participation in post-compulsory education and training: Report of the Australian Education Council Review Committee [Finn Review]*. Canberra: Australian Government Publishing Service.
- Fischer, G. H. (1995). Derivations of the Rasch model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 15-38). New York: Springer.
- Flanagan, J. C. (1951). Educational measurement. In E. F. Lindquist, *Educational measurement*. (pp. xix, 819-xix, 819). Washington DC: American Council on Education.
- Flórez Petour, M. T. (2015). Systems, ideologies and history: a three-dimensional absence in the study of assessment reform processes. *Assessment in Education: Principles, Policy & Practice*, 22(1), 3-26. <https://doi.org/10.1080/0969594X.2014.943153>
- Foley, D. (2003). Indigenous epistemology and Indigenous standpoint theory. *Social Alternatives*, 22(1), 44.

- Foster, J. B. (1983). Theories of Capitalist Transformation: Critical Notes on the Comparison of Marx and Schumpeter. *The Quarterly Journal of Economics*, 98(2), 327-331. <https://doi.org/10.2307/1885628>
- Foucault, M. (1988). *The care of the self: Volume 3 of the history of sexuality* (R. Hurley, Trans.). New York: Vintage Books. (Original work published 1984)
- Foucault, M. (1990). *The history of sexuality: Volume 1: An introduction* (R. Hurley, Trans.). New York: Vintage Books. (Original work published 1976)
- Foucault, M. (1990). *The use of pleasure: Volume 2 of the history of sexuality* (R. Hurley, Trans.). New York: Vintage Books. (Original work published 1984)
- Foucault, M. (1991). *Discipline and Punish: The Birth of the Prison* (A. Sheridan, Trans.). London: Penguin Books. (Original work published 1975)
- Foucault, M. (1991). Governmentality. In G. Burchell, C. Gordon, & P. Miller (Eds.), *The Foucault effect: Studies in governmentality: with two lectures and an interview with Michel Foucault*. Chicago: University of Chicago Press. (Original work published 1978)
- Foucault, M. (2002). *The order of things: An archaeology of the human sciences*. London: Routledge. (Original work published 1966)
- Fraser, N. (1990). Rethinking the public sphere: A contribution to the critique of actually existing democracy. *Social Text*, 25/26, 56-80. Available from <http://www.jstor.org/stable/466240>
- Fraser, N. (1995). What's critical about critical theory. In J. Meehan (Ed.), *Feminists read Habermas: Gendering the subject of discourse* (pp. 21-55). New York: Routledge.
- Fraser, N. (2013). *Fortunes of feminism: From state-managed capitalism to neoliberal crisis*. London: Verso Books.
- Fredricks, J. A., Filsecker, M., & Lawson, M. A. (2016). Student engagement, context, and adjustment: Addressing definitional, measurement, and methodological issues. *Learning and Instruction*, 43, 1-4. <https://doi.org/10.1016/j.learninstruc.2016.02.002>
- Freud, S. (1995). *The Freud reader* (P. Gay Ed.). London: Vintage.
- Friedman, M. (2002). *Capitalism and freedom* (40th Anniversary ed.). Chicago: University of Chicago Press. (Original work published 1962)

- Friedman, M. (2008). The methodology of positive economics. In D. M. Hausman (Ed.), *Essays in Positive Economics* (3rd ed., pp. 145-178). Cambridge UK: Cambridge University Press. (Original work published 1953)
- Friedman, M. (2008). *The optimum quantity of money*. New Brunswick NJ: Transaction Publishers. (Original work published 1969)
- Friedman, M. (2008). *Price theory*. New Brunswick NJ: Transaction Publishers. (Original work published 1962)
- Fukuyama, F. (1992). *The end of history and the last man*. London: Penguin Books.
- Gadamer, H.-G. (2008). *Philosophical hermeneutics* (D. E. Linge, Trans.). Berkeley: University of California Press. (Original work published 1976)
- Gadamer, H.-G. (2013). *Truth and method* (J. Weinsheimer & D. G. Marshall, Trans.). London: Bloomsbury. (Original work published 1975)
- Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1995). *Design patterns: Elements of reusable object-oriented software*. Boston: Addison-Wesley.
- Gandal, N., & Halaburda, H. (2014). *Competition in the cryptocurrency market* (Bank of Canada Working Paper No. 2014-33,). Ottawa: Bank of Canada.
- Gandhi, L. (1998). *Postcolonial theory: A critical introduction*. St Leonards: Allen & Unwin.
- Garner, W. R., Hake, H. W., & Eriksen, C. W. (1956). Operationism and the concept of perception. *Psychological Review*, 63(3), 149-159.
<https://doi.org/10.1037/h0042992>
- Garrett, D. (2015). *Hume*. London: Routledge.
- Gates, G. J. (2011). *How many people are lesbian, gay, bisexual, and transgender?* Los Angeles: The Williams Institute. Retrieved from www.law.ucla.edu/williamsinstitute
- Gebhardt, E., & Adams, R. J. (2007). The influence of equating methodology on reported trends in PISA. *Journal of Applied Measurement*, 8(3), 305-322.
- Gebser, J. (1985). *The Ever-present origin* (N. Barstad & A. Mickunas, Trans.). Athens OH: Ohio University Press.

- Gierl, M. J., & Lai, H. (2016). Automatic item generation. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds.), *The handbook of test development* (2 ed., pp. 410-429). New York: Routledge.
- Gillard, J. (2010). My School website launched [Press release]. Retrieved from <https://ministers.employment.gov.au/gillard/my-school-website-launched>
- Gillies, D. (2013). *Educational leadership and Michel Foucault*. London: Routledge.
- Gilligan, C. (1977). In a different voice: Women's conceptions of self and of morality. *Harvard Educational Review*, 47(4), 481-517. <https://doi.org/10.17763/haer.47.4.g6167429416hg510>
- Gillis, S., Polesel, J., & Wu, M. (2016). PISA Data: Raising concerns with its use in policy settings. *The Australian Educational Researcher*, 43(1), 131-146. <https://doi.org/10.1007/s13384-015-0183-2>
- Goldman, R., Heath, D., & Smith, S. L. (1991). Commodity feminism. *Critical Studies in Mass Communication*, 8(3), 333-351. <https://doi.org/10.1080/15295039109366801>
- Gonski, D. (2011). *Review of Funding for Schooling: Final Report*. Canberra: Department of Education.
- Gonski, D., Arcus, T., Boston, K., Gould, V., Johnson, W., O'Brien, L., . . . Roberts, M. (2018). *Through growth to achievement: Report of the review to achieve educational excellence in Australian schools*. Canberra: Commonwealth of Australia.
- Gordon Commission. (2013). *To assess, to teach, to learn: a vision for the future of assessment*. Retrieved from http://www.gordoncommission.org/rsc/pdfs/gordon_commission_technical_report.pdf
- Gorur, R. (2016). Seeing like PISA: A cautionary tale about the performativity of international assessments. *European Educational Research Journal*, 15(5), 598-616. <https://doi.org/10.1177/1474904116658299>
- Gorur, R., & Wu, M. (2015). Leaning too far? PISA, policy and Australia's 'top five' ambitions. *Discourse: Studies in the Cultural Politics of Education*, 36(5), 647-664. <https://doi.org/10.1080/01596306.2014.930020>
- Gramsci, A. (2014). *Selections from the prison notebooks of Antonio Gramsci* (Q. Hoare & G. N. Smith Eds.). New York: International Publishers. (Original work published 1971)

- Grek, S. (2009). Governing by numbers: the PISA 'effect' in Europe. *Journal of Education Policy*, 24(1), 23-37. <https://doi.org/10.1080/02680930802412669>
- Griffin, B., Bayl-Smith, P., & Hu, W. (2018). Predicting patterns of change and stability in student performance across a medical degree. *Medical education*, 52(4), 438-446. <https://doi.org/10.1111/medu.13508>
- Griffin, B., Harding, D. W., Wilson, I. G., & Yeomans, N. D. (2008). Does practice make perfect? The effect of coaching and retesting on selection tests used for admission to an Australian medical school. *The Medical Journal of Australia*, 189(5), 270-273.
- Griffin, P., & Care, E. (Eds.). (2015). *Assessment and teaching of 21st century skills: Methods and approaches*. Heidelberg: Springer Dordrecht. <https://doi.org/10.1007/978-94-017-9395-7>
- Griffin, P., Care, E., & McGaw, B. (Eds.). (2012). *Assessment and teaching of 21st century skills*. Dordrecht: Springer. <https://doi.org/10.1007/978-94-007-2324-5>
- Griffin, P., & Woods, K. (2006). *Evaluation of the pilot implementation of the Student@Centre Ultranet in Victorian schools*. Melbourne: University of Melbourne. Retrieved from <http://education.unimelb.edu.au/arc>
- Grisay, A. (2002). Translation and cultural appropriateness of the test and survey material. In R. J. Adams & M. Wu (Eds.), *PISA 2000 technical report* (pp. 57-70). Paris: OECD Publishing. <https://doi.org/10.1787/9789264199521-en>
- Grisay, A., de Jong, J. H. A. L., Gebhardt, E., Berezner, A., & Halleux-Monseur, B. (2007). Translation equivalence across PISA countries. *Journal of Applied Measurement*, 8(3), 249-266.
- Guyer, P., & Wood, A. W. (2016). (in Kant, I) Introduction to the Critique of Pure Reason (P. Guyer & A. W. Wood, Trans.) *Critique of pure reason*. New York: Cambridge University Press.
- Habermas, J. (1979). *Communication and the evolution of society* (T. McCarthy, Trans.). Boston: Beacon Press. (Original work published 1976)
- Habermas, J. (1981). Modernity versus Postmodernity (S. Ben-Habib, Trans.). *New German Critique*(22), 3-14. <https://doi.org/10.2307/487859>
- Habermas, J. (1985). *The theory of communicative action: Reason and the rationalization of society* (T. McCarthy, Trans. Vol. 1). Boston: Beacon Press. (Original work published 1981)

- Habermas, J. (1990). Jürgen Habermas: Morality, Society and Ethics: An interview with Torben Hviid Nielsen. *Acta Sociologica*, 33(2), 93-114.
<https://doi.org/10.1177/000169939003300201>
- Habermas, J. (1991). *The structural transformation of the public sphere: An inquiry into a category of bourgeois society* (T. Burger & F. Lawrence, Trans.). Cambridge MA: MIT press. (Original work published 1962)
- Habermas, J. (1992). *Postmetaphysical thinking: Philosophical essays* (W. M. Hohengarten, Trans.). Cambridge MA: MIT Press.
- Habermas, J. (1992). *The theory of communicative action: Lifeworld and system: a critique of functionalist reason* (T. McCarthy, Trans. Vol. 2). Boston: Beacon Press. (Original work published 1981)
- Habermas, J. (1994). *Justification and application: Remarks on discourse ethics* (C. Cronin, Trans.). Cambridge MA: MIT Press.
- Habermas, J. (1996). *Moral consciousness and communicative action* (C. Lenhardt & S. W. Nichol森, Trans.). Cambridge MA: MIT Press. (Original work published 1983)
- Habermas, J. (1997). Modernity: An unfinished project. In M. P. d'Entrèves & S. Benhabib (Eds.), *Habermas and the unfinished project of modernity: Critical essays on the philosophical discourse of modernity* (pp. 38-55). Cambridge MA: MIT Press.
- Habermas, J. (1998). *Between facts and norms: Contributions to a discourse theory of law and democracy* (W. Rehg, Trans.). Cambridge MA: MIT Press. (Original work published 1992)
- Habermas, J. (1998). *On the pragmatics of communication* (M. Cooke Ed.). Cambridge MA: MIT Press.
- Habermas, J. (1998). Toward a critique of the theory of meaning. In M. Cooke (Ed.), *On the pragmatics of communication* (pp. 277-306). Cambridge MA: MIT Press. (Original work published 1988)
- Habermas, J. (2000). *Inclusion of the other: Studies in political theory* (C. Cronin, Trans.). Cambridge, Massachusetts: MIT Press.
- Habermas, J. (2000). *The inclusion of the other: Studies in political theory* (C. Cronin & P. D. Greiff Eds.). Cambridge MA: MIT Press. (Original work published 1996)

- Habermas, J. (2001). *The postnational constellation: Political essays* (M. Pensky Ed.). Cambridge MA: MIT Press.
- Habermas, J. (2003). *Truth and justification* (B. Fultner, Trans.). Cambridge MA: MIT Press.
- Habermas, J. (2005). *Legitimation crisis* (T. McCarthy, Trans.). Boston: Beacon Press. (Original work published 1975)
- Habermas, J. (2007). *Theory and practice* (J. Viertel, Trans.). Cambridge UK: Polity. (Original work published 1973)
- Habermas, J. (2008). *Between naturalism and religion: Philosophical essays* (C. Cronin, Trans.). Cambridge UK: Polity.
- Habermas, J. (2015). *The future of human nature*. Cambridge UK: Polity. (Original work published 2003)
- Häkkinen, K., & Akrami, N. (2014). Ideology and climate change denial. *Personality and Individual Differences, 70*, 62-65.
<https://doi.org/10.1016/j.paid.2014.06.030>
- Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. New York: Routledge.
- Hambleton, R. K., & Pitoniak, M. J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 433-470). Westport, CT: American Council on Education and Praeger Publishers.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park CA: Sage publications.
- Hammer, M. (1990). Reengineering work: Don't automate, obliterate. *Harvard Business Review, 68*(4), 104-112. Available from <http://www.hbr.org>
- Hammer, M., & Champy, J. (1994). *Reengineering the corporation: A manifesto for business revolution*. New York: HarperBusiness.
- Hanushek, E. A. (2013). Economic growth in developing countries: The role of human capital. *Economics of Education Review, 37*, 204-212.
<https://doi.org/10.1016/j.econedurev.2013.04.005>

- Hanushek, E. A. (2016). Will more higher education improve economic growth? *Oxford Review of Economic Policy*, 32(4), 538-552.
<https://doi.org/10.1093/oxrep/grw025>
- Hanushek, E. A., & Ettema, E. (2017). Defining productivity in education: Issues and illustrations. *The American Economist*.
<https://doi.org/10.1177/0569434516688207>
- Hanushek, E. A., Schwerdt, G., Wiederhold, S., & Wößmann, L. (2013). *Returns to skills around the world: Evidence from PIAAC* (OECD Education Working Papers No. 101). Paris: OECD Publishing.
<https://doi.org/10.1787/5k3tsjqmvtq2-en>
- Hanushek, E. A., & Wößmann, L. (2010). *How much do educational outcomes matter in OECD countries?* (NBER Working Paper No. 16515). Cambridge MA: The National Bureau of Economic Research. Retrieved from
<http://www.nber.org/papers/w16515>
- Harcourt, B. E. (2007). *An answer to the question: 'What is poststructuralism?'* (Public Law Working Paper No. 156). Chicago: University of Chicago.
<https://doi.org/10.2139/ssrn.970348>
- Hardy, I., & Lewis, S. (2016). The 'doublethink' of data: educational performativity and the field of schooling practices. *British Journal of Sociology of Education*, 1-13.
<https://doi.org/10.1080/01425692.2016.1150155>
- Harlen, W. (2009). Improving assessment of learning and for learning. *Education 3-13*, 37(3), 247-257. <https://doi.org/10.1080/03004270802442334>
- Harlen, W., & James, M. (1997). Assessment and Learning: differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice*, 4(3), 365-379.
<https://doi.org/10.1080/0969594970040304>
- Harris, D. N., & Herrington, C. D. (2015). Editors' introduction: The use of teacher value-added measures in schools: New evidence, unanswered questions, and future prospects. *Educational Researcher*, 44(2), 71-76.
<https://doi.org/10.3102/0013189X15576142>
- Hart, R. (1992). *Children's participation: From tokenism to citizenship* (Innocenti Essays No. 4). Florence: United Nations Children's Fund. Retrieved from
https://www.unicef-irc.org/publications/pdf/childrens_participation.pdf
- Harvey, D. (1990). *The condition of postmodernity: An enquiry into the origins of cultural change*. Cambridge MA: Blackwell.

- Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Hattie, J. A. C., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Hausman, J. (2003). Sources of bias and solutions to bias in the consumer price index. *Journal of Economic Perspectives*, 17(1), 23-44. <https://doi.org/10.1257/089533003321164930>
- Hayek, F. A. (1980). *The fatal conceit: The errors of socialism* (W. W. Bartley Ed.). Chicago: University of Chicago Press. (Original work published 1988)
- Hayek, F. A. (2007). *The road to serfdom: Text and documents: The definitive edition* (B. Caldwell Ed.). London: The University of Chicago Press. (Original work published 1944)
- Hayek, F. A. (Ed.) (1963). *Collectivist economic planning*. London: Routledge & Kegan Paul. (Original work published 1935)
- Heath, J. (2011). System and lifeworld. In B. Fultner (Ed.), *Jürgen Habermas: Key concepts* (pp. 74-90). Durham: Acumen.
- Hegel, G. W. F. (1977). *Phenomenology of Spirit* (A. V. Miller, Trans.). Oxford: Oxford University Press. (Original work published 1807)
- Hegel, G. W. F. (1998). Science of logic. In S. Houlgate (Ed.), *The Hegel reader* (pp. 127-272). Malden: Blackwell Publishing. (Original work published 1816)
- Heidegger, M. (1996). *Being and Time: A Translation of Sein und Zeit* (J. Stambaugh, Trans.). Albany: State University of New York Press.
- Heiskala, R. (2007). Economy and society: From Parsons through Habermas to semiotic institutionalism. *Social Science Information*, 46(2), 243-272. <https://doi.org/10.1177/0539018407076648>
- Hildebrand, G. M. (1996). Redefining achievement. In P. F. Murphy & C. Gipps (Eds.), *Equity in the classroom: Towards effective pedagogy for girls and boys* (pp. 149-172). London: The Falmer Press.
- Hill, N. E., Castellino, D. R., Lansford, J. E., Nowlin, P., Dodge, K. A., Bates, J. E., & Pettit, G. S. (2004). Parent academic involvement as related to school behavior, achievement, and aspirations: Demographic variations across adolescence. *Child Development*, 75(5), 1491-1509. <https://doi.org/10.1111/j.1467-8624.2004.00753.x>

- Hinshelwood, R. D. (1987). *What happens in groups: Psychoanalysis, the individual and the community*. London: Free Association Books.
- Hofstede, G. (1994). Management scientists are human. *Management Science*, 40(1), 4-13. Available from <http://www.jstor.org/stable/2632841>
- Hogan, A. (2012). The changing nature of the Australian education policy field: The rise of edu-business. *Australian Association for Research in Education (NJ1)*. Available from <http://eric.ed.gov/?id=ED544480>
- Hogan, A. (2016a). NAPLAN and the role of edu-business: New governance, new privatisations and new partnerships in Australian education policy. *The Australian Educational Researcher*, 43(1), 93-110. <https://doi.org/10.1007/s13384-014-0162-z>
- Hogan, A. (2016b). # tellPearson: the activist 'public education' network. *Discourse: Studies in the Cultural Politics of Education*, 1-16. <https://doi.org/10.1080/01596306.2016.1269225>
- Hogan, A., Sellar, S., & Lingard, B. (2015). Network restructuring of global edu-business. In W. Au & J. J. Ferrare (Eds.), *Mapping corporate education reform. Power and policy networks in the neoliberal state* (pp. 43-64). New York: Routledge.
- Hogan, A., Sellar, S., & Lingard, B. (2016). Commercialising comparison: Pearson puts the TLC in soft capitalism. *Journal of Education Policy*, 31(3), 243-258. <https://doi.org/10.1080/02680939.2015.1112922>
- Holland, P. W., & Dorans, N. J. (2006). Linking and equating. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 187-220). Westport, CT: American Council on Education and Praeger Publishers.
- Holton, R. J., & Turner, B. S. (1986). *Talcott Parsons on economy and society*. London: Routledge.
- Howell, A. (2012). The silent voice in the NAPLAN debate: Exploring children's lived experiences of the tests. *Proceedings of Joint Australian Association for Research in Education and Asia-Pacific Educational Research Association Conference (AARE-APERA 2012) World Education Research Association (WERA) Focal Meeting*. Sydney, New South Wales. <https://eric.ed.gov/?id=ED544498>
- Howell, A. (2017). 'Because then you could never ever get a job!': children's constructions of NAPLAN as high-stakes. *Journal of Education Policy*, 32(5), 564-587. <https://doi.org/10.1080/02680939.2017.1305451>

- Huang, Y. (2014). *Pragmatics* (2nd ed.). Oxford: Oxford University Press.
- Hull, C. L. (1932). The goal-gradient hypothesis and maze learning. *Psychological Review*, 39(1), 25-43. <https://doi.org/10.1037/h0072640>
- Hume, D. (1985). *Treatise on human nature* (A. Selby-Bigge Ed.). Oxford: Oxford University Press. (Original work published 1739)
- Hunter-Doniger, T. (2017). The eugenics movement and its impact on art education in the United States. *Arts Education Policy Review*, 118(2), 83-92. <https://doi.org/10.1080/10632913.2015.1051256>
- Husserl, E. (1970). *The crisis of European sciences and transcendental phenomenology: An introduction to phenomenological philosophy* (C. David, Trans.). Evanston IL: Northwestern University Press. (Original work published 1954)
- IBAC. (2016). *Operation Dunham Investigation: Transcript of proceedings*. Transcripts from public hearing, Independent Broad-based Anti-corruption Commission. Melbourne. Retrieved 22 March 2017 from <http://www.ibac.vic.gov.au/investigating-corruption/public-examinations/operation-dunham>
- IBAC. (2017). *Operation Dunham: An investigation into the conduct of officers of the Department of Education and Training, including Darrell Fraser, in connection with the Ultramet project and related matters*. Melbourne: Independent Broad-based Anti-corruption Commission. Retrieved from <http://www.ibac.vic.gov.au/publications-and-resources/article/op-dunham-sr-jan-17>
- Illich, I. (1973). *Deschooling society*. Middlesex: Penguin Books. (Original work published 1970)
- Ingram, D. (2010). *Habermas: Introduction and analysis*. Ithaca NY: Cornell University Press.
- Irigaray, L. (1985). *This sex which is not one* (C. Porter, Trans.). New York: Cornell University Press.
- Irigaray, L. (1995). The question of the other (N. Guynn, Trans.). *Yale French Studies*(87), 7-19. Available from <http://www.jstor.org/stable/2930321>
- Jameson, F. (1984). Foreword. In J.-F. Lyotard, *The postmodern condition: A report on knowledge* (pp. vii-xxi). Minneapolis: University of Minnesota Press.

- Jameson, F. (1991). *Postmodernism, or, the cultural logic of late capitalism*. Durham: Duke University Press.
- Jarvis, C. H., Dickie, J., & Brown, G. (2013). Going mobile: perspectives on aligning learning and teaching in geography. *Journal of Geography in Higher Education*, 37(1), 76-91. <https://doi.org/10.1080/03098265.2012.694069>
- Jelinek, M. (1979). *Institutional innovation: A study of organizational learning systems*. New York: Praeger Publishers.
- Jessop, B. (2017). On academic capitalism. *Critical Policy Studies*, 12(1), 104-109. <https://doi.org/10.1080/19460171.2017.1403342>
- Jones, B. (1995). *Sleepers, wake!: Technology and the future of work* (New ed.). Melbourne: Oxford University Press. (Original work published 1982)
- Kane, M. T. (1990). *An argument-based approach to validation* (ACT Research Report Series No. ACT-RR-90-13). Iowa City IA: The American College Testing Program.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological bulletin*, 112(3), 527-535.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). Westport, CT: American Council on Education and Praeger Publishers.
- Kane, M. T. (2013). Validation as a pragmatic, scientific activity. *Journal of Educational Measurement*, 50(1), 115-122. <https://doi.org/doi:10.1111/jedm.12007>
- Kane, M. T. (2016a). Explicating validity. *Assessment in Education: Principles, Policy & Practice*, 23(2), 198-211. <https://doi.org/10.1080/0969594x.2015.1060192>
- Kane, M. T. (2016b). Validity as the evaluation of the claims based on test scores. *Assessment in Education: Principles, Policy & Practice*, 23(2), 309-311. <https://doi.org/10.1080/0969594x.2016.1156645>
- Kant, I. (1977). *Prolegomena to any future metaphysics* (J. W. Ellington & P. Carus, Trans. 2nd ed.). Indianapolis: Hackett Publishing. (Original work published 1772)
- Kant, I. (2002). *Groundwork for the metaphysics of morals* (A. W. Wood, Trans.). New Haven CT: Yale University Press. (Original work published 1785)

- Kant, I. (2004). *Critique of practical reason* (T. K. Abbott, Trans.). New York: Dover. (Original work published 1788)
- Kant, I. (2005). *Critique of judgement* (J. H. Bernard, Trans.). New York: Dover. (Original work published 1790)
- Kant, I. (2009). Lectures on pedagogy. In S. M. Cahn (Ed.), *Philosophy of education: The essential texts*. New York: Routledge. (Original work published 1803)
- Kant, I. (2016). *Critique of pure reason* (P. Guyer & A. W. Wood, Trans.). New York: Cambridge University Press. (Original work published 1781)
- Karimi, J., & Walter, Z. (2015). The Role of Dynamic Capabilities in Responding to Digital Disruption: A Factor-Based Study of the Newspaper Industry. *Journal of Management Information Systems*, 32(1), 39-81. <https://doi.org/10.1080/07421222.2015.1029380>
- Karpin, D. (1995). Enterprising nation: renewing Australia's managers to meet the challenges of the Asia-Pacific century [Karpin report]. *Canberra: Australian Government*.
- Katz, I. R. G., Joanne S. (2016). Computerising assessment: Impacts on education stakeholders. In G. T. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 472-489). New York: Routledge.
- Kautz, T., Heckman, J. J., Diris, R., ter Weel, B., & Borghans, L. (2014). *Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success* (OECD Education Working Papers No. 110). Paris: OECD Publishing. <https://doi.org/10.1787/5jxsr7vr78f7-en>
- Keddie, A. (2016). Children of the market: Performativity, neoliberal responsibilisation and the construction of student identities. *Oxford Review of Education*, 42(1), 108-122. <https://doi.org/10.1080/03054985.2016.1142865>
- Kegan, R. (1996). *The evolving self: Problem and process in human development*. Cambridge MA: Harvard University Press. (Original work published 1982)
- Kegan, R. (1997). *In over our heads: The mental demands of modern life*. Cambridge MA: Harvard University Press. (Original work published 1994)
- Kellner, D. (1992). *Critical theory, Marxism and modernity*. Baltimore: The Johns Hopkins University Press.
- Kellner, D. (2003). Toward a critical theory of education. *Democracy & Nature*, 9(1), 51-64. <https://doi.org/10.1080/1085566032000074940>

- Keynes, J. M. (1964). *General theory of employment, interest and money*. San Diego CA: Harvest Book. (Original work published 1953)
- Kirschner, P. A., & De Bruyckere, P. (2017). The myths of the digital native and the multitasker. *Teaching and Teacher Education*, 67, 135-142. <https://doi.org/10.1016/j.tate.2017.06.001>
- Klieme, E. (2016). *TIMSS 2015 and PISA 2015: How are they related on the country level ?* Frankfurt: German Institute for International Educational Research (DIPF). Retrieved from <https://pisa.dipf.de/de/pisa-aktuell/wie-timss-und-pisa-2015-zusammenhaengen>
- Kohl, G. O., Lengua, L. J., & McMahon, R. J. (2000). Parent involvement in school conceptualizing multiple dimensions and their relations with family and demographic risk factors. *Journal of School Psychology*, 38(6), 501-523. [https://doi.org/10.1016/S0022-4405\(00\)00050-9](https://doi.org/10.1016/S0022-4405(00)00050-9)
- Kohlberg, L. (1971). Stages of moral development as a basis for moral education. In C. M. Beck, B. S. Crittenden, & E. V. Sullivan (Eds.), *Moral education: Interdisciplinary approaches* (pp. 23-92). New York: Newman Press.
- Kohlberg, L., & Hersh, R. H. (1977). Moral development: A review of the theory. *Theory Into Practice*, 16(2), 53-59. <https://doi.org/10.1080/00405847709542675>
- Kompridis, N. (2006). *Critique and disclosure: Critical theory between past and future*. Cambridge MA: MIT Press.
- Koomen, M., & Zoanetti, N. (2018). Strategic planning tools for large-scale technology-based assessments. *Assessment in Education: Principles, Policy & Practice*, 25(2), 200-223. <https://doi.org/10.1080/0969594x.2016.1173013>
- Kotler, P. (1963). The use of mathematical models in marketing. *Journal of Marketing*, 27(4), 31-41. Available from <http://www.jstor.org/stable/1248643>
- Kotler, P. (1964). Marketing mix decisions for new products. *Journal of Marketing Research*, 1(1), 43-49. Available from <http://www.jstor.org/stable/3150319>
- Kotler, P. (1965). The competitive marketing simulator—A new management tool. *California Management Review*, 7(3), 49-60. <https://doi.org/10.2307/41165632>
- Kotler, P. (1972). A generic concept of marketing. *Journal of Marketing*, 36(2), 46-54. Available from <http://www.jstor.org/stable/1250977>
- Kotler, P., & Levy, S. J. (1969). Broadening the concept of marketing. *The Journal of Marketing*, 33(1), 10-15. Available from <http://www.jstor.org/stable/1248740>

- Kovas, Y., Tikhomirova, T., Selita, F., Tosto, M. G., & Malykh, S. (2016). How genetics can help education. *Behavioural Genetics for Education*, 1-23. https://doi.org/10.1057/9781137437327_1
- Kress, G. (2003). *Literacy in the New Media Age*. London: Routledge.
- Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. London: Routledge.
- Kuhn, T. (1970). *The structure of scientific revolutions* (Second Enlarged ed.). Chicago IL: The University of Chicago Press. (Original work published 1962)
- Lakoff, G., & Johnson, M. (2003). *Metaphors we live by*. Chicago: The University of Chicago Press. (Original work published 1980)
- Lamb, S., Jackson, J., Walstab, A., & Huo, S. (2015). *Educational opportunity in Australia 2015: Who succeeds and who misses out*. Melbourne: Mitchell Institute.
- Lambert, K., Wright, P. R., Currie, J., & Pascoe, R. (2015). Data-driven performativity: Neoliberalism's impact on drama education in Western Australian secondary schools. *Review of Education, Pedagogy, and Cultural Studies*, 37(5), 460-475. <https://doi.org/10.1080/10714413.2015.1091260>
- Laming, M. M. (2012). *The new inheritors: Transforming young people's expectations of university*. Rotterdam: Sense Publishers.
- Lane, S., Raymond, M. R., & Haladyna, T. M. (Eds.). (2016). *Handbook of test development*. New York: Routledge.
- Larochelle, M., Bednarz, N., & Garrison, J. (Eds.). (1998). *Constructivism and education*. Cambridge UK: Cambridge University Press.
- Larusson, J. A., & White, B. (Eds.). (2014). *Learning analytics: From research to practice*. New York: Springer.
- Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The attribute hierarchy method for cognitive assessment: A variation on Tatsuoka's rule-space approach. *Journal of Educational Measurement*, 41(3), 205-237. <https://doi.org/10.1111/j.1745-3984.2004.tb01163.x>
- Leitch, V. B. (1992). *Cultural criticism, literary theory, poststructuralism*. New York: Columbia University Press.

- Levy, F., & Murnane, R. J. (2004). *The new division of labor: How computers are creating the next job market*. Princeton: Princeton University Press.
- Lewandowsky, S., Oreskes, N., Risbey, J. S., Newell, B. R., & Smithson, M. (2015). Seepage: Climate change denial and its effect on the scientific community. *Global Environmental Change*, 33, 1-13.
<https://doi.org/10.1016/j.gloenvcha.2015.02.013>
- Lindquist, E. F. (Ed.) (1951). *Educational measurement*. Washington DC: American Council on Education.
- Lingard, B. (2011). Policy as numbers: Ac/counting for educational research. *The Australian Educational Researcher*, 38(4), 355-382.
<https://doi.org/10.1007/s13384-011-0041-9>
- Lingard, B., Sellar, S., Hogan, A., & Thompson, G. (2017). *Commercialisation in public schooling (CIPS)*. Sydney: News South Wales Teachers Federation.
- Linn, R. L. (Ed.) (1989). *Educational measurement* (3rd ed.). Washington DC: American Council on Education.
- Lipnevich, A. A., Berg, D. A. G., & Smith, J. K. (2016). Toward a model of student response to feedback. In G. T. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 169-185). New York: Routledge.
- Lischka, J. A. (2017). A Badge of Honor? *Journalism Studies*, 1-18.
<https://doi.org/10.1080/1461670x.2017.1375385>
- Lord, F. M., & Novick, M. R. (2008). *Statistical theories of mental test scores*. Oxford: Addison-Wesley. (Original work published 1968)
- Lowe, R. (1980). Eugenics and education: a note on the origins of the intelligence testing movement in England. *Educational Studies*, 6(1), 1-8.
<https://doi.org/10.1080/0305569800060101>
- Luhmann, N. (1992). What is communication. *Communications theory*, 2(3), 251-259.
<https://doi.org/10.1111/j.1468-2885.1992.tb00042.x>
- Luhmann, N. (1995). *Social systems* (J. Bednarz & D. Baecker, Trans.). Stanford CA: Stanford University Press. (Original work published 1984)
- Luhmann, N. (2016). System as difference. *Organization*, 13(1), 37-57.
<https://doi.org/10.1177/1350508406059638>

- Lyotard, J.-F. (1984). *The postmodern condition: A report on knowledge*. Minneapolis: University of Minnesota Press. (Original work published 1979)
- Lyotard, J.-F. (2015). *Libidinal economy* (I. H. Grant, Trans.). London: Bloomsbury. (Original work published 1974)
- Mangen, A., Walgermo, B. R., & Brønnick, K. (2013). Reading linear texts on paper versus computer screen: Effects on reading comprehension. *International Journal of Educational Research*, 58, 61-68.
<https://doi.org/10.1016/j.ijer.2012.12.002>
- Marginson, S. (1989). *Human capital theory and education policy* (Discussion Paper No. 3). Kensington NSW: Public Sector Research Centre, University of New South Wales Sydney.
- Marginson, S. (1993). *Education and public policy in Australia*. Cambridge: Cambridge University Press.
- Marginson, S. (1997a). *Educating Australia: Government, economy and citizen since 1960*. Cambridge UK: Cambridge University Press.
- Marginson, S. (1997b). *Markets in education*: Allen & Unwin Sydney.
- Markus, K. A., & Borsboom, D. (2013). *Frontiers of test validity theory: Measurement, causation, and meaning*. New York: Routledge.
- Martin, A. J., & Lazendic, G. (2018). Computer-adaptive testing: Implications for students' achievement, motivation, engagement, and subjective test experience. *Journal of Educational Psychology*, 110(1), 27-45.
<https://doi.org/10.1037/edu0000205>
- Martin, C. (2012). *Education in a post-metaphysical World: Rethinking educational policy and practice through Jürgen Habermas' discourse morality*. London: Bloomsbury Publishing.
- Martin, M. O., & Kelly, D. L. (1996). *TIMSS technical report volume I: Design and development*. Chestnut Hill, Massachusetts: International Association for the Evaluation of Educational Achievement.
- Marx, K. (2000). The Communist manifesto. In D. McLellan (Ed.), *Karl Marx: Selected writings* (pp. 246-272). Oxford: Oxford University Press. (Original work published 1848)
- Marx, K. (2000). Grundrisse. In D. McLellan (Ed.), *Karl Marx: Selected writings* (pp. 379-423). Oxford: Oxford University Press. (Original work published 1941)

- Marx, K. (2000). *Karl Marx: Selected writings* (D. McLellan Ed.). Oxford: Oxford University Press.
- Marx, K. (2000). Preface to a critique of political economy. In D. McLellan (Ed.), *Karl Marx: Selected writings* (pp. 424-428). Oxford: Oxford University Press. (Original work published 1859)
- Marx, K., & Engels, F. (2000). The German ideology. In D. McLellan (Ed.), *Karl Marx: Selected writings* (pp. 175-208). Oxford: Oxford University Press. (Original work published 1846)
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149-174. <https://doi.org/10.1007/BF02296272>
- Masters, G. N. (2002). *Fair and meaningful measures*. Camberwell: Australian Council for Educational Research.
- Masters, G. N. (2005, November 22). Classroom results undercut panic merchants' claims. *The Australian*.
- Masters, G. N. (2013). *Reforming educational assessment: Imperatives, principles and challenges*. Camberwell: Australian Council for Educational Research. Retrieved from <http://research.acer.edu.au/aer/12>
- Matters, G. (2015, 16–18 August). *Translating rich learning assessments into certified results and university selection devices*. Paper presented at the Learning assessments: Designing the future, Melbourne, Australia.
- Matters, G., & Masters, G. N. (2014). *Redesigning the secondary–tertiary interface: Queensland review of senior assessment and tertiary entrance*. Brisbane, Australia: Australian Council for Educational Research.
- Mayer, E. (1992). *Key competencies: Report of the Committee to advise the Australian Education Council and Ministers of Vocational Education, Employment and Training on employment-related key competencies for postcompulsory education and training [Mayer report]*. Canberra: Australian Education Council and Ministers of Vocational Education, Employment and Training.
- McCallin, R. C. (2016). Test administration. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds.), *The handbook of test development* (2 ed., pp. 567-584). New York: Routledge.
- McCarthy, T. (2005). Translator's introduction. In J. Habermas, *Legitimation crisis* (pp. vii-xxiv). Boston: Beacon Press. (Original work published 1975)

- McClure, K. R. (2016). Building the innovative and entrepreneurial university: An institutional case study of administrative academic capitalism. *The Journal of Higher Education*, 87(4), 516-543.
<https://doi.org/10.1080/00221546.2016.11777412>
- McCrae, B., & Stacey, K. (1997). Testing problem solving in a high-stakes environment. In E. Pehkonen (Ed.), *Use of open-ended problems in mathematics classroom: Research report 176* (pp. 34-48). Helsinki: University of Helsinki.
- McCrae, R. R., & Costa, P. T. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Personality and Individual Differences*, 36(3), 587-596.
[https://doi.org/10.1016/s0191-8869\(03\)00118-1](https://doi.org/10.1016/s0191-8869(03)00118-1)
- McCurry, D. (2010). Can machine scoring deal with broad and open writing tests as well as human readers? *Assessing Writing*, 15(2), 118-129.
<https://doi.org/10.1016/j.asw.2010.04.002>
- McCurry, D. (2017). *The art of item writing: Cross-curricular testing at ACER 1964 to 2014* (ACER Research Monograph No. 66). Camberwell: Australian Council for Educational Research. Retrieved from
http://research.acer.edu.au/acer_monographs/8
- McGorry, P., Bates, T., & Birchwood, M. (2013). Designing youth mental health services for the 21st century: Examples from Australia, Ireland and the UK. *The British Journal of Psychiatry*, 202(s54), s30-s35.
<https://doi.org/10.1192/bjp.bp.112.119214>
- McIntosh, P. (1986, October). *White privilege: Unpacking the invisible knapsack*. Paper presented at the American Educational Research Association, Boston. Retrieved from
<https://nationalseedproject.org/white-privilege-and-male-privilege>
- Meehan, J. (Ed.) (1995). *Feminists read Habermas: Gendering the subject of discourse*. New York: Routledge.
- Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46(4). <https://doi.org/10.1037/0022-006X.46.4.806>
- Mercer, A., Hay, M., Hodgson, W. C., Canny, B. J., & Puddey, I. B. (2018). The relative predictive value of undergraduate versus graduate selection tools in two Australian medical schools. *Med Teach*, 1-8.
<https://doi.org/10.1080/0142159X.2018.1426839>

- Merlino, J. (2017). Victoria not proceeding with NAPLAN online in 2017 [Press release]. Retrieved from <http://www.premier.vic.gov.au/victoria-not-proceeding-with-naplan-online-in-2017/>
- Merlino, J. (2018). Best NAPLAN results yet for Victorian kids [Press release]. Retrieved from <https://www.premier.vic.gov.au/best-naplan-results-yet-for-victorian-kids/>
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35(11), 1012-1027. <https://doi.org/10.1037/0003-066X.35.11.1012>
- Messick, S. (1988). *Consequences of test interpretation and use: The fusion of validity and values in psychological assessment* (ETS Research Report Series No. RR-98-48). Princeton, NJ: Educational Testing Service.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). Washington DC: American Council on Education.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13-23. Available from <http://www.jstor.org/stable/1176219>
- Meyer, H.-D. (2000). Taste formation in pluralistic societies: The role of rhetorics and institutions. *International Sociology*, 15(1), 33-56. <https://doi.org/10.1177/0268580900015001003>
- Milone, A. S., Cortese, A. M., Balestrieri, R. L., & Pittenger, A. L. (2017). The impact of proctored online exams on the educational experience. *Currents in Pharmacy Teaching and Learning*, 9(1), 108-114. <https://doi.org/10.1016/j.cptl.2016.08.037>
- Mintzberg, H. (1994). *The rise and fall of strategic planning: Reconceiving roles for planning, plans, planners*. New York: The Free Press.
- Mislevy, R. J. (2006). Cognitive psychology and educational assessment. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 257-305). Westport, CT: American Council on Education and Praeger Publishers.
- Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A brief introduction to evidence-centered design*. Unpublished Paper.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). Focus Article: On the Structure of Educational Assessments. *Measurement: Interdisciplinary Research & Perspective*, 1(1), 3-62. https://doi.org/10.1207/S15366359MEA0101_02

- Mislevy, R. J., Steinberg, L. S., Almond, R. G., Haertel, G. D., & Penuel, W. R. (2003). *Leverage points for improving educational assessment* (Principled Assessment Designs for Inquiry (PADI) Technical Report No. 2). Menlo Park: SRI International. Retrieved from <https://padi.sri.com/publications.html>
- Mockler, N. (2013). Reporting the 'education revolution': MySchool. edu. au in the print media. *Discourse: Studies in the Cultural Politics of Education*, 34(1), 1-16. <https://doi.org/10.1080/01596306.2012.698860>
- Moore, M. H. (1995). *Creating public value: Strategic management in government*. Cambridge MA: Harvard University Press.
- Moore, M. H. (2013). *Recognizing public value*. Cambridge MA: Harvard University Press.
- Moran, D. (2000). *Introduction to phenomenology*. London: Routledge.
- Morgan, G. (1997). *Images of Organization*. California: SAGE Publications.
- Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research*, 62(3), 229-258. <https://doi.org/10.3102/00346543062003229>
- Moss, P. A. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice*, 22(4), 13-25. <https://doi.org/10.1111/j.1745-3992.2003.tb00140.x>
- Moss, P. A. (2016). Shifting the focus of validity for test use. *Assessment in Education: Principles, Policy & Practice*, 23(2), 236-251. <https://doi.org/10.1080/0969594x.2015.1072085>
- Mouzelis, N. (1992). Social and system integration: Habermas' view. *The British Journal of Sociology*, 43(2), 267-288. Available from <http://www.jstor.org/stable/591468>
- Mouzelis, N. (1997). Social and system integration: Lockwood, Habermas, Giddens. *Sociology*, 31(1), 111-119. <https://doi.org/10.1177/0038038597031001008>
- Murdock, T. B., Stephens, J. M., & Grotewiel, M. M. (2016). Student dishonesty in the face of assessment. In G. T. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 186-203). New York: Routledge.
- Murphy, M., & Fleming, T. (2010). Communication, deliberation, reason: An introduction to Habermas. In M. Murphy & T. Fleming (Eds.), *Habermas, critical theory and education* (pp. 3-16). New York: Routledge.

- Navarrete, T., Santos, P., Hernández-Leo, D., & Blat, J. (2011). QTIMaps: A model to enable web maps in assessment. *Educational Technology & Society*, 14(3), 203-217. Available from <http://www.jstor.org/stable/jeductechsoci.14.3.203>
- Newmann, F., Brandt, R., & Wiggins, G. (1998). An exchange of views on "semantics, psychometrics, and assessment reform: A close look at 'authentic' assessments". *Educational Researcher*, 27(6), 19-22. Available from <http://www.jstor.org/stable/1176091>
- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, 14(2), 149-170. <https://doi.org/10.1080/09695940701478321>
- Newton, P. E. (2012). Validity, purpose and the recycling of results from educational assessments. In J. Gardner (Ed.), *Assessment and Learning* (2nd ed., pp. 264-276). London: SAGE Publications.
- Newton, P. E., & Baird, J.-A. (2016a). The great validity debate. *Assessment in Education: Principles, Policy & Practice*, 23(2), 173-177. <https://doi.org/10.1080/0969594x.2016.1172871>
- Newton, P. E., & Baird, J.-A. (Eds.). Validity [Special issue]. *Assessment in Education: Principles, Policy & Practice*, 23(2). <https://doi.org/10.1080/0969594x.2016.1172871>
- Newton, P. E., & Shaw, S. D. (2014). *Validity in educational and psychological assessment*. London: SAGE.
- Nisbet, R. (2012). *The present age: Progress and anarchy in modern America*. Indianapolis: Liberty Fund. (Original work published 1988)
- Nisbet, R. (2017). *Metaphor and history: The western idea of social development*. London: Routledge. (Original work published 1969)
- Noddings, N. (2010). Moral education and caring. *School Field*, 8(2), 145-151. <https://doi.org/10.1177/1477878510368617>
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- O'Loughlin, J., Ni Chróinín, D., & O'Grady, D. (2013). Digital video: The impact on children's learning experiences in primary physical education. *European Physical Education Review*, 19(2), 165-182. <https://doi.org/10.1177/1356336X13486050>

- Oakeshott, M. (1991). *Rationalism in Politics and Other Essays*. Indianapolis: Liberty Fund. (Original work published 1962)
- OECD. (2000). *Measuring student knowledge and skills: The PISA 2000 assessment of reading, mathematical and scientific literacy. Education and skills*. PARIS: OECD Publishing. <https://doi.org/10.1787/9789264181564-en>
- OECD. (2001). *Knowledge and skills for life: First results from PISA 2000*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264195905-en>
- OECD. (2003). *The PISA 2003 Assessment Framework: Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264101739-en>
- OECD. (2005a). *PISA 2003 data analysis manual: SPSS®users*. Paris: OECD Publishing.
- OECD. (2005b). *PISA 2003 technical report*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264010543-en>
- OECD. (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264026407-en>
- OECD. (2007). *PISA 2006 science competencies for tomorrow's world: Volume 1 – analysis*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264040014-en>
- OECD. (2009a). *PISA 2006 technical report*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264048096-en>
- OECD. (2009b). *PISA data analysis manual: SPSS® second edition*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264056275-en>
- OECD. (2010a). *PISA 2009 assessment framework: Key competencies in reading, mathematics and science*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264062658-en>
- OECD. (2010b). *PISA computer-based assessment of student skills in science*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264082038-en>
- OECD. (2011). *The protection of children online: Risks faced by children online and policies to protect them* (OECD Digital Economy Papers No. 179). Paris: OECD Publishing. <https://doi.org/10.1787/5kgcjf71pl28-en>

- OECD. (2012). *PISA 2009 technical report*. Paris: OECD Publishing.
<https://doi.org/10.1787/9789264167872-en>
- OECD. (2013a). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264190511-en>
- OECD. (2013b). *PISA 2015 draft collaborative problem solving framework*. Retrieved from <https://www.oecd.org/pisa/pisaproducts/pisa2015draftframeworks.htm>
- OECD. (2013c). *Synergies for better learning: An international perspective on evaluation and assessment*. Paris: OECD Publishing.
<https://doi.org/10.1787/9789264190658-en>
- OECD. (2014). *PISA 2012 technical report*. Paris: OECD Publishing.
- OECD. (2016a). *Education at a glance 2016: OECD indicators*. Paris: OECD Publishing. <https://doi.org/10.1787/19991487>
- OECD. (2016b). *PISA 2015 results (Volume I): Excellence and equity in education*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264266490-en>
- OECD. (2016c). *PISA 2015 results (Volume II): Policies and practices for successful schools*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264267510-en>
- OECD. (2017a). *OECD Economic Surveys: Australia 2017*. Paris: OECD Publishing.
https://doi.org/10.1787/eco_surveys-aus-2017-en
- OECD. (2017b). *PISA 2015 assessment and analytical framework: Science, reading, mathematics, financial literacy and collaborative problem solving* (Revised ed.). Paris: OECD Publishing. <https://doi.org/10.1787/9789264281820-en>
- OECD. (2017c). *PISA 2015 results (Volume III): Student's well-being*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264273856-en>
- OECD. (2017d). *PISA 2015 results (Volume IV): Students' financial literacy*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264270282-en>
- OECD. (2017e). *PISA 2015 technical report*. Paris: OECD Publishing.
- OECD. (2017f). *PISA 2015 technical standards PISA 2015 Technical Report* (pp. 438-454). Paris: OECD Publishing.

- OECD and PISA tests are damaging education worldwide – academics. (2014, 6 May). *The Guardian*. Retrieved from <https://www.theguardian.com/education/2014/may/06/oecd-pisa-tests-damaging-education-academics>
- Olson, K. (2011). Deliberative democracy. In B. Fultner (Ed.), *Jürgen Habermas: Key concepts* (pp. 140-155). Durham: Acumen.
- Pajnik, M. (2006). Feminist reflections on Habermas's communicative action: The need for an inclusive political theory. *European Journal of Social Theory*, 9(3), 385-404. <https://doi.org/10.1177/1368431006065719>
- Papert, S. (1993). *Mindstorms: Children, computers, and powerful ideas* (2nd ed.). New York: Basic Books.
- Parsons, T. (1963a). On the concept of influence. *The Public Opinion Quarterly*, 27(1). Available from <http://www.jstor.org/stable/2747290>
- Parsons, T. (1963b). On the concept of political power. *Proceedings of the American Philosophical Society*, 107(3), 232-262. Available from <http://www.jstor.org/stable/985582>
- Parsons, T. (1968a). *The structure of social action: A study in social theory with special reference to a group of recent European writers* (Vol. 2). New York: The Free Press. (Original work published 1937)
- Parsons, T. (1968b). *The structure of social action: A study in social theory with special reference to a group of recent European writers* (Vol. 1). New York: The Free Press. (Original work published 1937)
- Parsons, T. (2012). *The social system*. New Orleans: Quid Pro Books. (Original work published 1951)
- Parsons, T., & Platt, G. M. (1970). Age, social structure, and socialization in higher education. *Sociology of Education*, 43(1), 1-37. <https://doi.org/10.2307/2112057>
- Parsons, T., & Shils, E. A. (Eds.). (1962). *Toward a general theory of action: Theoretical foundations for the social sciences*. New York: Harper & Row. (Original work published 1951)
- Parsons, T., & Smelser, N. (2005). *Economy and society: A study in the integration of economic and social theory*. New York: Routledge. (Original work published 1956)

- Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex* (G. V. Anrep, Trans.). New York: Oxford University Press.
- Pearson, D. D. R. (2009). *Literacy and numeracy achievement*. Melbourne: Victorian Auditor-General's Office. Retrieved from <http://www.audit.vic.gov.au/>
- Pearson, D. D. R. (2012). *Learning technologies in government schools*. Melbourne: Victorian Auditor-General's Office. Retrieved from <http://www.audit.vic.gov.au/>
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington DC: The National Academies.
- Penney, D., Jones, A., Newhouse, C. P., & Cambell, A. (2012). Developing a digital assessment in senior secondary physical education. *Physical Education & Sport Pedagogy*, 17(4), 383-410. <https://doi.org/10.1080/17408989.2011.582490>
- Perelman, L. (2018). *Towards a new NAPLAN: Testing to the teaching*. Surry Hills: NSW Teachers Federation. Retrieved from www.nswtf.org.au
- Perie, M., & Huff, K. L. (2016). Determining content and cognitive demand for achievement tests. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds.), *Handbook of test development* (2nd ed., pp. 119-143). New York: Routledge.
- Peters, R. S. (1973). *Authority, responsibility and education* (3rd ed.). London: George Allen & Unwin. (Original work published 1959)
- Phillips, D. C. (1995). The good, the bad, and the ugly: The many faces of constructivism. *Educational Researcher*, 24(7), 5-12. Available from <http://www.jstor.org/stable/1177059>
- Phillips, S. E., & Camara, W. J. (2006). Legal and ethical issues. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 733-755). Westport, CT: American Council on Education and Praeger Publishers.
- Piaget, J. (1973). *Structuralism* (C. Maschler, Trans.). London: Routledge and Kegan Paul. (Original work published 1968)
- Piaget, J. (1997). *The moral judgment of the child* (M. Gabain, Trans.). New York: The Free Press. (Original work published 1932)
- Piaget, J. (2001). *The psychology of intelligence* (M. Piercy & D. E. Berlyne, Trans.). London: Routledge. (Original work published 1947)

- Pinar, W., Reynolds, W. M., Slattery, P., & Taubman, P. M. (2008). *Understanding curriculum: An introduction to the study of historical and contemporary curriculum discourses*. New York: Peter Lang Publishing.
- Piolat, A., Roussey, J.-Y., & Thunin, O. (1997). Effects of screen presentation on text reading and revising. *International Journal of Human-Computer Studies*, 47(4), 565-589. <https://doi.org/10.1006/ijhc.1997.0145>
- Plato. (1963). *The republic*. Middlesex: Penguin Books.
- Plato. (1993). *Sophist* (N. P. White, Trans.). Indianapolis: Hackett Publishing.
- Polesel, J., Rice, S., & Dulfer, N. (2013). The impact of high-stakes testing on curriculum and pedagogy: a teacher perspective from Australia. *Journal of Education Policy*, 29(5), 640-657. <https://doi.org/10.1080/02680939.2013.865082>
- Pollak, R. A., & Wales, T. J. (1992). *Demand system specification and estimation*. New York: Oxford University Press.
- Poole, P., Shulruf, B., Rudland, J., & Wilkinson, T. (2012). Comparison of UMAT scores and GPA in prediction of performance in medical school: a national study. *Medical education*, 46(2), 163-171. <https://doi.org/10.1111/j.1365-2923.2011.04078.x>
- Popper, K. (2002). *The open society and its enemies*. London: Routledge. (Original work published 1945)
- Postman, N. (1993). *Technopoly: The surrender of culture to technology*. New York: Vintage Books.
- Postman, N. (1996). *The end of education: Redefining the value of school*. New York: Vintage Books.
- Powell, J. L. (2012). *Baudrillard and postmodernism*. New York: Nova Science Publishers.
- Prager, I. G., Moran, G., & Sanchez, J. (1997). Job analysis of felony assistant public defenders: The most important tasks and most useful knowledge, Skills, and Abilities. *Psychology, Crime & Law*, 3(1), 37-49. <https://doi.org/10.1080/10683169608409793>
- Prensky, M. (2001). Digital natives, digital immigrants part 1. *On the Horizon*, 9(5), 1-6. <https://doi.org/10.1108/10748120110424816>

- Puddey, I. B., Mercer, A., Andrich, D., & Styles, I. (2014). Practice effects in medical school entrance testing with the undergraduate medicine and health sciences admission test (UMAT). *BMC Medical Education*, 14(48).
<https://doi.org/10.1186/1472-6920-14-48>
- Pyne, C. (2014). \$24.7 million to transform NAPLAN to online testing [Press release]. Retrieved from <https://ministers.education.gov.au/pyne/247-million-transform-naplan-online-testing>
- Rahim, E. (2009). Marx and Schumpeter: A Comparison of their Theories of Development. *Review of Political Economy*, 21(1), 51-83.
<https://doi.org/10.1080/09538250802516982>
- Rahimi, M., Halse, C., & Blackmore, J. (2017). Transnational secondary schooling and im/mobile international students. *The Australian Educational Researcher*, 44(3), 299-321. <https://doi.org/10.1007/s13384-017-0235-x>
- Ramsay, G. (2017). NAPLAN online roll out delayed in ACT [Press release]. Retrieved from http://www.cmd.act.gov.au/open_government/inform/act_government_media_releases/gordon-ramsay-mla-media-releases/2017/naplan-online-roll-out-delayed-in-act
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: MESA PRESS. (Original work published 1960)
- Rawls, J. (1999). *A theory of justice* (Revised ed.). Cambridge MA: Belknap Press. (Original work published 1971)
- Reay, D., & Wiliam, D. (1999). 'I' ll be a nothing': structure, agency and the construction of identity through assessment. *British Educational Research Journal*, 25(3), 343-354. <https://doi.org/10.1080/0141192990250305>
- Reckase, M. D. (2017). *A tale of two models: Sources of confusion in achievement testing* (Research Report No. RR-17-44). Princeton, NJ: Educational Testing Service. <https://doi.org/10.1002/ets2.12171>
- Rehg, W. (1998). Translator's introduction (W. Rehg, Trans.). In J. Habermas, *Between facts and norms: Contributions to a discourse theory of law and democracy* (pp. i-xxxvii). Cambridge MA: MIT Press. (Original work published 1992)
- Rehg, W. (2011). Discourse ethics. In B. Fultner (Ed.), *Jürgen Habermas: Key concepts* (pp. 115-139). Durham: Acumen.

- Reserve Bank Act 1959* (Austl.). Available from <https://www.legislation.gov.au/Details/C2015C00201>
- Rickwood, D. J., Telford, N. R., Parker, A. G., Tanti, C. J., & McGorry, P. D. (2014). headspace - Australia's innovation in youth mental health: who are the clients and why are they presenting? *The Medical journal of Australia*, 200(2), 108-111. Available from <https://www.ncbi.nlm.nih.gov/pubmed/24484115>
- Riconscente, M. M., Mislevy, R. J., & Corrigan, S. (2016). Evidence-centered design. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds.), *Handbook of test development* (2nd ed., pp. 40-63). New York: Routledge.
- Rioch, M. J. (1970). The work of Wilfred Bion on groups. *Psychiatry*, 33(1), 56-66. <https://doi.org/10.1080/00332747.1970.11023613>
- Ritchhart, R., & Perkins, D. N. (2005). Learning to think: The challenges of teaching thinking. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 775-802). Cambridge UK: Cambridge University Press.
- Rizmal, Z. (2017, June 2). Anti-safe schools parents' group targets Victoria's Deputy Premier. *The Age*. Retrieved from <http://www.theage.com.au/victoria/antisafe-schools-parents-group-targets-victorias-deputy-premier-20170601-gwi2i2.html>
- Robinson, N. (2018, August 27). This year's NAPLAN results of 'very limited use' and should be discarded, US experts argue. *Australian Broadcasting Corporation*. Retrieved from <http://www.abc.net.au/news/2018-08-27/naplan-testing-report-says-results-should-be-discarded/10160596>
- Rodriguez, M. C. (2003). Construct equivalence of multiple-choice and constructed-response items: A random effects synthesis of correlations. *Journal of Educational Measurement*, 40(2), 163-184. Available from <http://www.jstor.org/stable/1435344>
- Rodwell, G. (2016). Safe schools: Challenges to moral panic and risk society theories: The politics of national school educational policy. *Discourse Journal of Educational Research*, 3(2), 16. Available from <http://www.resjournals.org/IJER/Abstract/2016/July/Rodwell.htm>
- Rogers, C. R. (1959). Significant learning: In therapy and in education. *Educational Leadership*, 16(4).
- Rogers, C. R. (2003). *Client centred therapy: Its current practice, implications and theory*: Little, Brown Book Group. (Original work published 1951)

- Rorty, R. (1984). Habermas and Lyotard on post-modernity. *Praxis International*, 4(1), 32-44. Available from <https://www.ceeol.com/search/journal-detail?id=615>
- Rorty, R. (1995). Habermas, Derrida, and the functions of philosophy. *Revue Internationale de Philosophie*, 49(194(4)), 437-459. Available from <http://www.jstor.org/stable/23954536>
- Rousseau, J.-J. (1979). *Emile, or on education* (A. Bloom, Trans.). New York: Basic Books. (Original work published 1762)
- Roussos, L. A., DiBello, L. V., Stout, W., Hartz, S. M., Henson, R. A., & Templin, J. (2007). The fusion model skills diagnosis system. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 275-318). New York: Cambridge University Press.
- Rudd, K., & Gillard, J. (2008). *Quality Education: The case for an education revolution in our schools*. Canberra: Commonwealth of Australia.
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic Measurement: Theory, Methods, and Applications*. New York: Guilford Press.
- Rupp, A. A., & Templin, J. L. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement: Interdisciplinary Research and Perspective*, 6(4), 219-262. <https://doi.org/10.1080/15366360802490866>
- Ryan, A. M., Mcfarland, L., & Baron, H. (1999). An international look at selection practices: Nation and culture as explanations for variability in practice. *Personnel Psychology*, 62(2), 359-392. <https://doi.org/10.1111/j.1744-6570.1999.tb00165.x>
- Ryle, G. (2000). *The concept of mind*. London: Penguin Books. (Original work published 1949)
- Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education*, 34(2), 159-179. <https://doi.org/10.1080/02602930801956059>
- Said, E. W. (1994). *Orientalism* (25th Anniversary ed.). New York: Vintage Books. (Original work published 1978)
- Sandene, B., Horkay, N., Bennett, R. E., Allen, N., Braswell, J., Kaplan, B., & Oranje, A. (2005). *Online assessment in mathematics and writing: Reports from the NAEP technology-based assessment project, research and development series*. Washington DC: U.S. Government Printing Office.

- Saurette, P. (2005). *Habermas's Kantian imperative: Humiliation, common sense, politics*. Toronto: University of Toronto Press.
- Savin-Williams, R. C., & Diamond, L. M. (2000). Sexual identity trajectories among sexual-minority youths: Gender comparisons. *Archives of Sexual Behavior*, 29(6), 607-627. <https://doi.org/10.1023/a:1002058505138>
- Scalise, K., & Gifford, B. (2006). Computer-based assessment in e-Learning: A framework for constructing "intermediate constraint" questions and tasks for technology platforms. *The Journal of Technology, Learning, and Assessment*, 4(6). Available from <http://www.jtla.org>
- Schiappa, E. (2013). *Protagoras and logos: A study in Greek philosophy and rhetoric* (2nd ed.). Columbia: University of South Carolina Press.
- Schibeci, R. A. (1984). Attitudes to Science: an update. *Studies in Science Education*, 11(1), 26-59. <https://doi.org/10.1080/03057268408559913>
- Schilt, K., & Bratter, J. (2015). From multiracial to transgender? Assessing attitudes toward expanding gender options on the US Census. *Transgender Studies Quarterly*, 2(1), 77-100. <https://doi.org/10.1215/23289252-2848895>
- Schmeiser, C. B., & Welch, C. J. (2006). Test development. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 307-353). Westport, CT: American Council on Education and Praeger Publishers.
- Schmidt, E., & Cohen, J. (2010). The digital disruption: Connectivity and the diffusion of power. *Foreign Affairs*, 89(6), 75-85. Available from <https://www.jstor.org/stable/20788718>
- Schneewind, J. B. (2002). Why study Kant's ethics (A. W. Wood, Trans.). In A. W. Wood (Ed.), *Groundwork for the metaphysics of morals* (pp. 83-91). New Haven CT: Yale University Press.
- Schulz, K. F., Altman, D. G., Moher, D., & Group, C. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *British Medical Journal*, 340(7748), 698-702. <https://doi.org/10.1136/bmj.c332>
- Schulz, W., Ainley, J., Fraillon, J., Losito, B., & Agrusti, G. (2016). *IEA international civic and citizenship education study 2016 assessment framework*. Switzerland: Springer Open.
- Schumpeter, J. A. (2008). *Capitalism, socialism and democracy* (3rd ed.). New York: Harper Perennial Modern Thought. (Original work published 1942)

- Seeratan, K., & Mislevy, R. J. (2008). *Design patterns for assessing internal knowledge representations* (Principled Assessment Designs for Inquiry (PADI) Technical Report No. 22). Menlo Park: SRI International. Retrieved from <https://padi.sri.com/publications.html>
- Shakeshaft, N. G., Trzaskowski, M., McMillan, A., Rimfeld, K., Krapohl, E., Haworth, C. M., . . . Plomin, R. (2013). Strong genetic influence on a UK nationwide test of educational achievement at the end of compulsory education at age 16. *PLoS One*, 8(12), e80341. <https://doi.org/10.1371/journal.pone.0080341>
- Shepard, L. A. (2000). *The role of classroom assessment in teaching and learning* (CSE Technical Report No. 517). Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Retrieved from <http://cresst.org/publications/cresst-publication-2889/>
- Shepard, L. A. (2006). Classroom Assessment. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 623-646). Westport, CT: American Council on Education and Praeger Publishers.
- Shepard, L. A. (2016). Evaluating test validity: Reprise and progress. *Assessment in Education: Principles, Policy & Practice*, 23(2), 268-280. <https://doi.org/10.1080/0969594x.2016.1141168>
- Shermis, M. D., Burstein, J. C., Brew, C., Higgins, D., & Zechner, K. (2016). Recent innovations in matching scoring of student- and test taker-written and-spoken responses. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds.), *Handbook of test development* (2 ed., pp. 335-354). New York: Routledge.
- Shi, Q., Pavey, E. S., & Carter, R. E. (2012). Bonferroni-based correction factor for multiple, correlated endpoints. *Pharm Stat*, 11(4), 300-309. <https://doi.org/10.1002/pst.1514>
- Shine, K. (2016). Reporting the Gonski review of school funding. *Australian Journalism Review*, 38(1).
- Sireci, S. G., & Zenisky, A. L. (2016). Computerised innovative item formats: Achievement and credentialing. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds.), *Handbook of test development* (2nd ed., pp. 313-334). New York: Routledge.
- Skinner, B. F. (1938). *The behaviour of organisms: An experimental analysis*. New York: Appleton-Century-Crofts.

- Skinner, B. F. (1957). The experimental analysis of behaviour. *American Scientist*, 45(4), 343-371. Available from <http://www.jstor.org/stable/27826953>
- Smith, A. (1993). *An inquiry into the nature and causes of the wealth of nations* (K. Sutherland Ed.). Oxford: Oxford University Press. (Original work published 1776)
- Snow, R. E., & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). Washington DC: American Council on Education.
- Solomon, Y., & Lewin, C. (2016). Measuring ‘progress’: Performativity as both driver and constraint in school innovation. *Journal of Education Policy*, 31(2), 226-238. <https://doi.org/10.1080/02680939.2015.1062147>
- Spaull, A. (1999). The end of the state school system? Education and the Kennett government. In B. Costar & N. Economou (Eds.), *The Kennett Revolution: Victorian politics in the 1990s*. Sydney, New South Wales: University of New South Wales Press.
- Spivak, G. C. (2010). Can the subaltern speak? In R. C. Morris (Ed.), *Can the subaltern speak?: Reflections on the history of an idea* (pp. 21-78). New York: Columbia University Press. (Original work published 1985)
- Stacey, K. (2001). Trends in researching and teaching problem solving in school mathematics in Australia: 1997-2000. In E. Pehkonen (Ed.), *Problem Solving around the World* (pp. 45-53). Finland: University of Turku.
- Stahl, T. (2013). Habermas and the project of immanent critique. *Constellations*, 20(4), 533-552. <https://doi.org/10.1111/1467-8675.12057>
- Stanovich, K. E. (1988). Cautions regarding Bonferroni Corrections: A response to Davies & Goldsmith. *Reading Research Quarterly*, 23(3), 380-381. Available from <http://www.jstor.org/stable/748049>
- Stedman Jones, D. (2012). *Masters of the universe: Hayek, Friedman, and the birth of neoliberal politics*. Princeton, NJ: Princeton University Press.
- Stevens, M. J., & Campion, M. A. (1994). The knowledge, skill, and ability requirements for teamwork: Implications for human resource management. *Journal of Management*, 20(2), 503-530. <https://doi.org/10.1177/014920639402000210>
- Stillman, G. (2007). Implementation case study: Sustaining curriculum change. In W. Blum, P. L. Galbraith, H.-W. Henn, & M. Niss (Eds.), *Modelling and*

applications in mathematics education: The 14th ICMI study (pp. 497-502). New York: Springer.

- Szypszak, C. (2015). Socratic Method for the Right Reasons and in the Right Way: Lessons from Teaching Legal Analysis Beyond the American Law School. *Journal of Political Science Education*, 11(3), 358-369. <https://doi.org/10.1080/15512169.2015.1047106>
- Tan, E. (2014). Human capital theory: A holistic criticism. *Review of Educational Research*, 84(3), 411-445. <https://doi.org/10.3102/0034654314532696>
- Tatsuoka, K. K. (2009). *Cognitive Assessment: An Introduction to the Rule Space Method*. New York: Routledge.
- Taylor, F. W. (1998). *The principles of scientific management*. New York: Dover Publications. (Original work published 1911)
- Thompson, N. A., & Weiss, D. J. (2011). A framework for the development of computerized adaptive tests. *Practical Assessment, Research & Evaluation*, 16(1), 1-9. Available from <http://pareonline.net/getvn.asp?v=16&n=1>
- Thomson, S., De Bortoli, L., & Buckley, S. (2013). *PISA 2012: How Australia measures up*. Camberwell: Australian Council for Educational Research. Retrieved from www.acer.edu.au/ozpisa/reports/
- Thomson, S., De Bortoli, L., & Underwood, C. (2016). *PISA 2015: A first look at Australia's results*. Camberwell: Australian Council for Educational Research. Retrieved from www.acer.edu.au/ozpisa/reports/
- Thomson, S., De Bortoli, L., & Underwood, C. (2017). *PISA 2015: Reporting Australia's results*. Camberwell: Australian Council for Educational Research. Retrieved from www.acer.edu.au/ozpisa/reports/
- Thorndike, R. L. (Ed.) (1971). *Educational measurement* (2nd ed.). Washington DC: American Council on Education.
- Thurlow, M. L., & Kopriva, R. J. (2015). Advancing accessibility and accommodations in content assessments for students with disabilities and English learners. *Review of Research in Education*, 39(1), 331-369. <https://doi.org/10.3102/0091732x14556076>
- Tierney, R. D., & Koch, M. J. (2016). Privacy in classroom assessment. In G. T. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 267-283). New York: Routledge.

- Toulmin, S. E. (1970). *An examination of the place of reason in ethics*. London: Cambridge University Press. (Original work published 1950)
- Toulmin, S. E. (2003). *The uses of argument* (Updated ed.). New York: Cambridge university press. (Original work published 1958)
- Triandis, H. C. (2001). Individualism-collectivism and personality. *Journal of Personality*, 69(6), 907-924. <https://doi.org/10.1111/1467-6494.696169>
- Turner, R., & Stacey, K. (2015). Werner Blum's contribution to PISA mathematics. In G. Kaiser & H.-W. Henn (Eds.), *Werner Blum und seine Beiträge zum Modellieren im Mathematikunterricht* (pp. 297-308): Springer Spektrum. https://doi.org/10.1007/978-3-658-09532-1_23
- U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research, & Center for Biologics Evaluation and Research. (1998). *Guidance for industry: E9 Statistical Principles for Clinical Trials*. Rockville MD: Office of Training and Communications Division of Drug Information. Retrieved from <https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm073137.pdf>
- van Der Linden, W. J. (2016). Optimal test assembly. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds.), *The handbook of test development* (2 ed., pp. 507-530). New York: Routledge.
- VCAA. (n.d-a). Be on a VCE examination development panel. Retrieved 19 May 2017 from <http://www.vcaa.vic.edu.au/Pages/vce/exams/exampanel.aspx>
- VCAA. (n.d-b). Past examinations and examination reports. Retrieved 19 May 2017 from <http://www.vcaa.vic.edu.au/Pages/vce/exams/examsassessreports.aspx>
- von Davier, A. (Ed.) (2011). *Statistical models for test equating, scaling, and linking*. New York: Springer.
- von Davier, M., Yamamoto, K., Shin, H. J., Chen, H., Khorramdel, L., Weeks, J., . . . Kandathil, M. (2019). Evaluating item response theory linking and model fit for data from PISA 2000–2012. *Assessment in Education: Principles, Policy & Practice*, 1-23. <https://doi.org/10.1080/0969594x.2019.1586642>
- von Mises, L. (2007). *Bureaucracy* (B. B. Greaves Ed.). Indianapolis: Liberty Fund. (Original work published 1944)
- VTAC. (2016). *ABC of scaling 2016*. Melbourne: Victorian Tertiary Admissions Centre. Retrieved from <http://www.vtac.edu.au>

- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman Eds.). Cambridge MA: Harvard University Press.
- Wajcman, J. (1991). Patriarchy, technology, and conceptions of skill. *Work and Occupations*, 18(1), 29-45. <https://doi.org/10.1177/0730888491018001002>
- Wajcman, J. (2004). *TechnoFeminism*. Cambridge UK: Polity.
- Watt, A. (1975). Transcendental arguments and moral principles. *The Philosophical Quarterly*, 25(98), 40-57. <https://doi.org/10.2307/2217951>
- Weber, M. (1964). *The theory of social and economic organization* (A. M. Henderson & T. Parsons, Trans.). New York: The Free Press. (Original work published 1915)
- Weber, M. (2009a). Bureaucracy (H. H. Gerth & C. W. Mills, Trans.). In H. H. Gerth & C. W. Mills (Eds.), *From Max Weber: Essays in sociology* (pp. 196-244). London: Routledge. (Original work published 1946)
- Weber, M. (2009b). *From Max Weber: Essays in sociology* (H. H. Gerth & C. W. Mills, Trans. H. H. Gerth & C. W. Mills Eds.). London: Routledge. (Original work published 1946)
- Weber, M. (2011). *The Protestant ethic and the spirit of capitalism* (S. Kalberg, Trans. Revised 1920 ed.). New York: Oxford University Press. (Original work published 1920)
- Weir, A. (1995). Toward a model of self-identity: Habermas and Kristeva. In J. Meehan (Ed.), *Feminists read Habermas: Gendering the subject of discourse* (pp. 263-282). New York: Routledge.
- Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21(4), 361-375. <https://doi.org/10.1111/j.1745-3984.1984.tb01040.x>
- Wendler, C. L., & Walker, M. E. (2016). Practical issues in designing and maintaining multiple test forms. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds.), *Handbook of test development* (2nd ed., pp. 433-449). New York: Routledge.
- Wesolowski, B. C., Wind, S. A., & Engelhard, G. (2015). Rater fairness in music performance assessment: Evaluating model-data fit and differential rater functioning. *Musicae Scientiae*, 19(2), 147-170. <https://doi.org/10.1177/1029864915589014>

- West, E. G. (2000). *An alternative to existing Australian research and teaching models : the Japangka teaching and research paradigm ; an Australian Aboriginal model*. (Unpublished doctoral thesis) Southern Cross University, Lismore, NSW.
- Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 70(9), 703-713. <https://doi.org/10.1177/003172171109200721>
- William, D. (2008). International comparisons and sensitivity to instruction. *Assessment in Education: Principles, Policy & Practice*, 15(3), 253-257. <https://doi.org/10.1080/09695940802417426>
- William, D. (2010). What counts as evidence of educational achievement? The role of constructs in the pursuit of equity in assessment. *Review of Research in Education*, 34(1), 254-284. <https://doi.org/10.3102/0091732x09351544>
- William, D. (2016). Foreword. In D. Didau, *What if everything you knew about education was wrong?* (pp. vii-x). Carmarthen: Crown House Publishing.
- Willms, D. (2003). *Student engagement at school: A sense of belonging and participation: Results from PISA 2000*. Paris: OECD Publishing.
- Wilson, M. R., & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education*, 13(2), 181-208. https://doi.org/10.1207/S15324818AME1302_4
- Wise, S. L., & Kingsbury, G. G. (2016). Modeling student test-taking motivation in the context of an adaptive achievement test. *Journal of Educational Measurement*, 53(1), 86-105. <https://doi.org/10.1111/jedm.12102>
- Wise, S. L., & Smith, L. F. (2016). The validity of assessment when students don't give good effort. In G. T. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 204-220). New York: Routledge.
- Wittgenstein, L. (2009). *Philosophical investigations* (G. E. M. Anscombe, P. M. S. Hacker, & J. Schulte, Trans.). West Sussex: Wiley-Blackwell. (Original work published 1953)
- Wittig, M. (1993). One is not born a woman. In H. Abelove, M. A. Barale, & D. M. Halperin (Eds.), *The lesbian and gay studies reader* (pp. 103-109). New York: Routledge.
- World Medical Association. (2013). *The World Medical Association Declaration of Helsinki - ethical principles for medical research involving human subjects*. Retrieved from www.wma.net

- Wren, D. A. (1994). *The Evolution of Management Thought* (4 ed.). New York: John Wiley & Sons.
- Wright, B. D. (1980). Foreword. In G. Rasch, *Probabilistic models for some intelligence and attainment tests* (pp. ix-xix). Chicago: MESA PRESS.
- Wright, B. D. (1997). A history of social science measurement. *Educational Measurement*, 16(4), 33-45. <https://doi.org/10.1111/j.1745-3992.1997.tb00606.x>
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. Chicago: MESA PRESS.
- Wu, M. (2009). A comparison of PISA and TIMSS 2003 achievement results in mathematics. *PROSPECTS*, 39, 33-46. <https://doi.org/10.1007/s11125-009-9109-y>
- Wu, M. (2014). Evidence-based policy making in education. *International Journal of Contemporary Educational Research*, 1(1), 1-8.
- Wu, M. (2016). What national testing data can tell us. In B. Lingard, G. Thompson, & S. Sellar (Eds.), *National testing in schools: An Australian assessment*. London: Routledge.
- Wu, M., Adams, R. J., & Wilson, M. R. (1997). ConQuest: Multi-Aspect Test Software [Computer software]. Camberwell: Australian Council for Educational Research.
- Wu, M., & Hornsby, D. (2014). Inappropriate uses of NAPLAN results. *Practically Primary*, 19(2), 16-17.
- Wu, M., Tam, H. P., & Jen, T.-H. (2017). *Educational measurement for applied researchers: Theory into practice*. Singapore: Springer. <https://doi.org/10.1007/978-981-10-3302-5>
- Wunderlich, D. (1979). *Foundations of linguistics* (R. Lass, Trans.). Cambridge UK: Cambridge University Press
- Wyatt-Smith, C., Klenowski, V., & Gunn, S. (2010). The centrality of teachers' judgement practice in assessment: a study of standards in moderation. *Assessment in Education: Principles, Policy & Practice*, 17(1), 59-75. <https://doi.org/10.1080/09695940903565610>
- Yen, W. M., & Fitzpatrick, A. R. (2006). Item Response Theory. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 111-153). Westport, CT: American Council on Education and Praeger Publishers.

- Zahavi, D. (2005). *Subjectivity and selfhood: Investigating the first-person perspective*. Cambridge MA: MIT Press.
- Zieky, M. J. (2012). So much has changed: An historical overview of setting cut scores. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed.). New York: Routledge.
- Zieky, M. J. (2016). Developing fair tests. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds.), *The handbook of test development* (2 ed., pp. 81-99). New York: Routledge.
- Zum, C. (2011). Discourse theory of law. In B. Fultner (Ed.), *Jürgen Habermas: Key concepts* (pp. 156-173). Durham: Acumen.
- Zumbo, B. D., & Hubley, A. M. (2016). Bringing consequences and side effects of testing and assessment to the foreground. *Assessment in Education: Principles, Policy & Practice*, 23(2), 299-303.
<https://doi.org/10.1080/0969594X.2016.1141169>