Deliberative democracy in educational assessment validation

Paper presented at AERA Conference 2019

April 5-9, 2019, Toronto, Canada
Session Title - Exploring Issues in Large-scale Assessments
Marten Koomen
College of Education, Victorian University - Melbourne

Abstract
This paper addresses the use of large-scale educational assessment data by professionals such as teachers, administrators and policy-makers. It does so by grounding approaches to educational assessment validation in principles of deliberative democracy. This is achieved by drawing on approaches to argument used in both educational assessment validation and deliberative democracy. A shared interest in validity across sociology and educational assessment allows educational assessment to be framed as having a centrally reported intended interpretation and an actual interpretation made in local contexts. This paper first details the foundational concepts for deliberative democracy in educational assessment validation, before illustrating how these concepts apply to the Programme for International Student Assessment (PISA). The discussion critiques aspects of the PISA methodology from the perspective of validation through deliberative democracy.

Introduction
This paper addresses a research priority identified by Moss (2016) to better conceptualise interpretation of educational assessment data in local contexts by professionals including teachers, administrators and policy-makers. This research priority is addressed by grounding Kane’s (2006) approach to validation into concepts of discourse ethics and deliberative democracy developed by Habermas (1996, 1998). This provides an enhanced framing for validation that includes intended interpretations reported by large-scale assessment programs as well as actual interpretations made by professionals in local contexts as described Moss (2016).1

1 O’Leary, Hattie, and Griffin (2017) make a later but similar framing of intended interpretation and actual interpretation independently to Moss (2016), this later framing is not pursued here.
The validity of large-scale educational assessment is increasingly being challenged. There are critiques describing an inordinate focus on the numerical component of educational assessment (Biesta, 2015; Gillis, Polesel, & Wu, 2016; Gorur & Wu, 2015; Grek, 2009; Lingard, 2011). There are also critiques around the general design and use of large-scale educational assessment, with several leading academics writing an open letter to the director of the Programme for International Student Assessment (PISA) ("OECD and PISA tests are damaging education worldwide – academics," 2014). Framing educational assessment through deliberative democracy allows these issues to be cast in terms of legitimate educational assessment design and use.

The method used in this paper follows the tradition of rational reconstruction (Carnap, 1962, p. 3; Habermas, 1979, p. 13; Wunderlich, 1979, pp. 169-172). Rational reconstruction focuses on explicating general theories and linking theories in a coherent way. Rational reconstruction requires reference to clear case examples to stabilise meaning. It is towards this requirement that this paper explores the PISA which is administered by the Organisation for Economic Cooperation and Development (OECD). Passing reference is also made to the Trends in International Mathematics and Science Study (TIMSS). The PISA is explored to illustrate how the concepts identified in this paper relate to contemporary practice, and to illustrate the reconstructive nature of the theory described in this paper.

This paper only claims to provide an overview of a general theory, and no claim to completeness is made.

**Foundational concepts**

Several concepts are used to frame the intended interpretation and actual interpretation proposed by Moss (2016). First is Toulmin’s (2003) approach to argument which links Kane’s (2006) argument-based approach to validation to Habermas’ (1996, 1998) approach to discourse ethics and deliberative democracy. The discourses of justification and application are integral to Habermas’ (1998) concept of deliberative democracy and are also explored (Habermas, 1994, 2003). Habermas’ (1992) conceptualisation of symbolic media is then used to conceptualise educational assessment data as symbolic media comprising a linguistic component and a numerical component. How these concepts manifest in contemporary practices of the PISA is then explored in the following section.
Toulmin’s (2003) approach to argument is used by both Kane (2006) and Habermas (1996) to ground their respective theories, and the approach is also used by Mislevy, Steinberg, and Almond (2003) to structure evidentiary arguments in educational assessment. Toulmin (2003) is used here to link approaches to educational assessment validation to discourse ethics and deliberative democracy.

A central distinction made by Toulmin (2003) is one between analytic and substantial argument. Analytic arguments are largely tautological as conclusions are contained within the argument. For example, “John is Jill’s son, all Jill’s sons are carpenters, so John is a carpenter” is an analytic argument where the evidence for the conclusion is contained in the argument. Educational assessment provides similar examples of analytic arguments such as “Country A is rank x, Country B is rank x+n, Country A must improve to catch up to Country B”. Conclusions from this type of argument are self-contained and do not require recourse to further evidence or reasoning. These types of arguments simply require regard for the information contained in the argument.

Further examples of analytic arguments are those that emerge from data analysis, syllogistic logic, and semantic analysis of expressions. These types of arguments are self-contained and closed as information for the argument is contained in the argument. For example, that Singapore is at the top of the PISA 2015 scale is an analytic argument, as once the argument is located in the PISA data, the conclusion can be found in that data through analysis without recourse to other evidence (OECD, 2016a, p. 67).

Substantial arguments address higher order predicates such as what is true, what is good, and what is right. A substantial argument might involve a claim that Singapore has the best education system for teaching science. It is for these types of arguments that additional data and warrants might be brought to bear to support the claim, as well as arguments and evidence that may rebut the claim. It is through argument drawing on various data and warrants that may lead to the claim being accepted, reject, or accepted on a qualified basis. In substantial arguments of this sort, additional evidence to that of the PISA data might brought to bear which might include data on the PISA methodology, data from TIMSS, or data pertaining to characteristics of Singapore’s education system.

Toulmin’s (2003) approach to substantial argument is illustrated in the Figure 1 below. Figure 1 illustrates how a claim, such as Singapore having the best education system for teaching
science, might be based on the PISA data, and that this claim may need to be supported by warrants and other data. The claim may also face challenge and rebuttal.

Following Toulmin (2003), a distinction can also made between empirical and ethical use of data. *Empirical use of data* is consistent with analytic argument where data is used to frame empirical arguments based on analysis. An example of empirical use of data is found in work of Hanushek and Ettema (2017) that explores productivity of education systems. The argument mounted by Hanushek and Ettema (2017) involves comparing analyses from educational assessment data as output measures from education with analyses from economic input measures. Hanushek and Ettema (2017) do not engage with higher order predicates such as if the data they employ is good, true, or fit for purpose. Hanushek and Ettema (2017) do not engage in substantial argument but only empirical argument.

*Ethical use of data* is found in the work Gorur and Wu (2015) who critique the use of numerical ranks from the PISA in Australian educational policy formulation. The notion of ethics is used here in the sense of Messick (1989) who argues that consideration of social consequences is part of validity. Gorur and Wu (2015) compare the PISA test content with local curriculum and are therefore making substantial argument by bringing further evidence to bear on claims based on the PISA data. Gorur and Wu (2015) are also making an ethical argument as it addresses imperatives around the social consequences of public policy. Gorur and Wu (2015) address what is right and what is good in Australian education and are not concerned with the kind of empirical arguments found in the analyses of Hanushek and Ettema (2017).

*Kane’s approach to validation*

Kane’s (2006) approach to educational assessment validation is grounded in Toulmin’s (2003) approach to substantial argument. Kane (2006) describes validation as involving two arguments. The first argument involves interpretation of data, the second argument involves
the validity of that interpretation. This paper seeks to extend the notion of the second interpretation to one of actual interpretation made in local contexts which draws on additional data from the local context.

Kane (2006, p. 29) considers interpretive arguments as arguments that cannot be “proven”. Instead, interpretive arguments are judged on their soundness in terms of clarity, coherence and plausibility. Here Kane’s (2006) approach is consistent with that of Habermas (1998) who finds that “the only thing that counts is the compelling force of the better argument based on the relevant information” (p. 103). This is to suggest that substantial argument in educational assessment validation cannot be proven in an empirical or analytic sense and instead need to be judged on the nature and quality of the argument itself.

Habermas, discourse ethics and deliberative democracy

While Toulmin (2003) is a common antecedent to both the work of Kane (2006) and Habermas (1996), Habermas (1996) identifies issues with Toulmin’s (2003) approach, issues that this paper argues also applies to Kane’s (2006) approach to validation.

Kane’s (2006) approach to validation is couched in terms of experts in centralised contexts. The approach is presaged on access to various data and information on test constructs, test content, and other supporting data. This information is generally not available outside of assessment agencies such as the OECD. This requires validation to be considered separately for central and local contexts. In a similar way, Habermas (1996) finds Toulmin’s (2003) approach to argument insufficiently normatively anchored and incapable of providing impartial consideration of validity claims. Habermas (1996) addresses this limitation in Toulmin’s (2003) approach by anchoring validity arguments in a principle of universalisation and a principle of discourse. These principles can also be used to extend Kane’s (2006) approach to include Moss’ (2016) actual interpretation in local contexts.

Habermas’ (1996) principle of universalisation requires the consequences of a claim to be acceptable to all affected, and Habermas’ (1996) principle of discourse requires all those affected by a claim to have an opportunity to participate in discussion. When these principles are applied to Kane’s (2006) approach to validation the perspectives of all those affected by an interpretation come into view, as does Moss’ (2016) actual interpretation in local contexts.

The principles of universalisation and discourse applies at an idealised level of citizen and Habermas (1998) extends these principles through the concept of deliberative democracy. For Habermas (1998, p. 298), successful deliberative process does not depend solely on the
participation of citizenry, but upon institutionalised procedures for the interplay between formally and informally developed public opinion. Habermas (1998, p. 367) describes the essence of deliberative democracy as will-formation through government bodies and civil bodies spontaneously emerging in society. In the context of education and educational assessment, formal bodies might include groups sponsored by government education departments and related agencies. Spontaneous bodies might include parent groups, subject-matter groups, trade unions groups, and so forth. Habermas (1998) argues that it is the quality of discourse across these bodies that drive the legitimacy of government decisions, laws and policies.

An important aspect of deliberative democracy pertinent to the present discussion are discourses of justification and discourses of application (Habermas, 1994, 1998). When considered in the context of intended and actual interpretation proposed by Moss (2016), discourses of justification are associated with intended interpretation reported by assessment agencies such as the OECD through the PISA. In this context, the reporting of results from assessment programs such as the PISA can be considered as intended interpretations that need to be justified with supporting arguments and information. In contrast, discourses of application are associated with the interpretations of PISA results in participating countries which may draw on data from the local context in making actual interpretations, as do Gorur and Wu (2015) for Australia.

There is also an emphasis on reciprocity between discourses of justification and discourses of application in Habermas’ (1994, 1998, 2003) formulation of deliberative democracy. Discourses of justification, such as those associated with intended interpretation reported by assessment programs such as the PISA, are for universal application and are largely indifferent to the cultural lives of citizens. Discourses of application are associated with actual interpretation made in local contexts and which considers the sociocultural context of citizens to which an intended interpretation might apply. In terms of reciprocity, whenever unintended consequences are observed in discourses of application, the central context needs to be responsive to these effects and recalibrate intended interpretation. The civil and spontaneous bodies described in Habermas’ (1998) concept of deliberative democracy are important here for identifying unintended consequences in local contexts emerging from centrally reported intended interpretation.
Conventional or deliberative democratic focus

This rudimentary explication of foundational concepts for deliberative democracy in educational assessment validation allows Moss’ (2016) framing of actual interpretation and intended interpretation to be sociologically situated. This is done here by comparing the conventional focus of validation (Figure 2) with a deliberative democratic approach described later.

Moss (2016) describes intended interpretation as the conventional focus of educational assessment validation as illustrated in Figure 2. Figure 2 locates validation in a central context where argument-based validation is not structurally connected to local contexts and isolated from those affected by an intended interpretation. This illustrates Habermas’ (1996) concern for Toulmin’s (2003) approach to argument, where unwarranted consensus may emerge in central contexts isolated from those in local contexts affected by the intended interpretation. Further, while the validation illustrated in Figure 2 addresses ethics consistent with Messick (1989) this consideration of ethics is only open to those in the central context, with those in the local context largely making empirical use of data. While this empirical focus may not always be the case in practice in local contexts, the conventional focus of validation nevertheless locates ethical reasoning in the central context.

Figure 3 provides an alternative framing for validation based on the work of Moss (2016) and Habermas (1994, 1998).
An approach to educational assessment validation focused on deliberative democracy as illustrated in Figure 3 includes both intended interpretation and actual interpretation as the focus of validation. This framing is not antagonistic to the argument-based approach to validation explicated by Kane (2006). What Figure 3 illustrates is that validation through deliberative democracy locates the validation process described by Kane (2006) in two spatiotemporally removed contexts. Further, Figure 3 illustrates that the types of validity arguments, such as those related to construct and content validity, are shared by the central and local contexts.

The types of data available in central and local contexts is likely to be different however. The central context is likely to have data available to inform arguments related to construct validity, content validity, predictive validity, and so forth. These data are likely to have a validity horizon reflecting the central context. In contrast, while the data in the local context informs similar arguments, the types of data used in the local context might be different. For example, in a local context arguments on construct validity might address the relevance of the assessed construct to the local context. This is illustrated by the work of Gorur and Wu (2015) who explore the relationship of the PISA test questions with the local curriculum in Australia. Validity arguments in local contexts might draw on further data from their context, and might involve triangulation with data from other centralised contexts, a theme explored by Wu (2009) and Klieme (2016) when comparing the results from TIMSS and the PISA. In this way, the validation in central and local contexts are similar, with validity arguments based on different data.
One major difference in the availability of data between central and local contexts pertains to test developer intentions. Kane (2016) observes that test developers are able to propose interpretations for test scores and “have considerable latitude in choosing what to include” (p. 209). Information about test developer intentions is generally not readily available in local contexts, particularly when the reported intended interpretation only involves numerical data. It is to this point that Habermas’ (1994, 2003) discourses of justification provides a useful framing. Mainly, that intended interpretations reported from central contexts and assessment agencies need to be supported by justifying information to support discourses of application in local contexts. This theme will be returned to later.

**Educational assessment as symbolic media**

The symbolic media between the spatiotemporally distinct intended interpretation and actual interpretation also needs to be conceptualised. The field of sociology provides a range of conceptualisations of symbolic media including that of Parsons (1963a, 1963b), Luhmann (1992, 1995) and Habermas (1992) (Chernilo, 2002). Habermas’ (1992) approach is used here.

Habermas (1992) provides a useful framing for symbolic media that mediates between the intended and actual interpretation through the term “delinguistified media” (p. 155). The term *delinguistified* alerts to fact that educational assessment has both a numerical component to indicate an achievement level and a linguistic component that communicates meaning. Habermas (1992) describes three forms of symbolic media for coordinating society. First is *ordinary language*, which is used to mount validity arguments, such validity arguments in central and local contexts illustrated in Figure 3. Second is *generalised communication* which is used to communicate complex meaning, such as communicating educational assessment data from a central context to a local context. An important characteristic explored by Habermas (1992) is that generalised communication remains open to validity arguments when applied in a local context and is not suitable for empirical use. This approach is consistent with Moss’ (2016) proposed approach illustrated in Figure 3 that shows that when an intended interpretation is reported its use in a local context remains open to validity argument. The third form of symbolic media is *steering media* which Habermas (1992) describes as abstract, delinguistified, and for empirical use. Habermas (1992) describes money and administrative power as steering media which are both used empirically in exchange without regard to meaning. Empirical use of educational assessment in local contexts is illustrated in Figure 2. This kind of empirical use is often found in discourses involving human capital theory which combines measures of educational assessment with measures of money to model educational
efficiency and its relationship to economic growth (Becker, 1993; Hanushek, 2016; Hanushek & Ettema, 2017). In this context, educational assessment as symbolic media becomes “delinguistified” through being combined in ratios with the abstract steering media of money.

The linguistic component of educational assessment as symbolic media encompasses many aspects. Cronbach (1971, p. 461) considers a brief descriptive title an important part of communicating meaning. Another important aspect is where the linguistic meaning of a test is invested. Classical test theory (Lord & Novick, 2008) invests meaning in test questions and students taking a test (Hambleton, Swaminathan, & Rogers, 1991, p. 7). In contrast, item response theory invests meaning in an explicit latent variable (Wu, Tam, & Jen, 2017, p. 21), in a way that Wright (1980, p. ix) describes as transcending test questions and the measuring instrument. When meaning is invested in a latent variable, how that meaning is conceptualised may also be contested, an issue not further pursued here (for example, see Borsboom, Mellenbergh, & van Heerden, 2003).

The numerical component of educational assessment is generally reported through one of five levels of measurement (Lord & Novick, 2008, pp. 20-21). However, levels of measurement only apply to linear measures such as those provided by classical test theory and item response theory (Hambleton et al., 1991). There are also non-linear measures such as those emerging from nomological networks (Cronbach & Meehl, 1955) and Bayesian models (Almond, Mislevy, Steinberg, Yan, & Williamson, 2015).

Only a preliminary sketch has been made here of the numerical and linguistic components of educational assessment. In summary, the linguistic component pertains to the meaning of test content and the numerical component pertains to a measure describing a student’s relationship to that content. This is sufficient for the argument made here.

Habermas (1992) develops an extensive theory around the notion of delinguistified media and its relationship to colonisation of the lifeworld. Colonisation refers to the process where generalised forms of communication are increasingly used in an empirical fashion without recourse to validity arguments. In this sense, colonisation refers to the process where the linguistic component of educational assessment data shears from its numerical component so that only the numerical component is used in educational discourse. This process is referred to by others as the seduction of numbers, policy by numbers, the tyranny of numbers or similar (Ball, 2015; Biesta, 2015; Gillis et al., 2016; Gorur & Wu, 2015; Grek, 2009; Lingard, 2011). The theory of colonisation is not further explored here, and for the present discussion it is
sufficient to note that educational assessment as symbolic media has a numerical component that can shear off from its meaningful linguistic component.

**Deliberative democratic validation in contemporary practice**

Consistent with the method of rational reconstruction, the foundational concepts are now explored through the PISA to provide clear case support that the identified foundational concepts are found in contemporary practice. However, no claim to completeness is made, nor is any claim made regarding the efficacy of the documented practices. Some issues emerging from these practices and related critique are provided in the following discussion section.

**Deliberative democratic validation in the PISA**

The basic architecture for validation through deliberative democracy in the PISA is illustrated in Figure 4. The key elements addressed are the central processes for validation, the reporting of data and supporting justifications, and the process of local validation of the actual interpretation. How the PISA caters for reciprocity is also addressed.

![Figure 4 – Deliberative democracy in the PISA](image)

**Central validation for intended interpretation**

The validity of the intended interpretations reported by the PISA is established through a complex structure of political and expert groups centred around the OECD. The PISA Governing Board provides the overarching framework for the PISA. Contractors are responsible for the design and implementation of the surveys within this framework. The OECD Secretariat is responsible for the day-to-day management of the PISA and fosters consensus across countries and acts as an interlocutor between the PISA Governing Board and Contractors (OECD, 2017c, pp. 26-27). The Contractor is also responsible for convening subject matter expert groups who assist in the validation of the assessment frameworks and the
items and instruments used for the PISA (OECD, 2017c, pp. 43-44). A technical expert group is also convened to assure the quality of data by way of further validation before it is included in data sets and reported (OECD, 2017c, p. 114).

**Justifying arguments for intended interpretations**

The PISA reports central intended interpretations through interpretive reports and tables such as the table showing Singapore as achieving the highest mean score for science in PISA 2015 (OECD, 2016a, p. 67). These tables are supplemented with other data and interpretations that address matters such as equity in education systems. The initial interpretive reports are supplemented by other volumes in the form of thematic reports addressing matters such as school policy, school practices, and student well-being (OECD, 2016b, 2017b). Further justificatory material is provided through the assessment and analytical frameworks that detail the basis for item development for the PISA (OECD, 2017a). Assessment frameworks are particularly useful for actual interpretation in local contexts and provide a basis for exploring how the content assessed in the PISA might correspond to the local curriculum and activity in the local economy. The technical report provides further justifying arguments for the methodology employed by the PISA (OECD, 2017c). The databases and codebooks are also available for secondary analyses (www.oecd.org/pisa/data/2015database/).

One further important form of justification is the public availability of test material. Issues related to the availability of test material will be addressed in the discussion.

**Validation for actual interpretation in the local context**

The PISA makes its data available for secondary analyses and provides support for individual countries to undertake secondary analyses (OECD, 2009). Further, actual interpretation for local jurisdictional contexts are reported through reports produced within participating countries as well as by the OECD (see, for example, Santiago, Donaldson, Herman, & Shewbridge, 2011; Thomson, De Bortoli, & Buckley, 2013; Thomson, De Bortoli, & Underwood, 2017).

Further, in Australia for example, the PISA data is used to support a range of policy arguments. This form of actual interpretation can be through government sponsored reports addressing matters such as the educational performance of Australia’s Indigenous community, Australia’s school funding policy, and Australian participation in science, technology engineering, and mathematics (Australian Government, 2018; Finkel, 2018; Gonski, 2011; Gonski et al., 2018). In addition to government reports, independent think tanks also use the PISA data to make
political arguments (Buckingham, 2016; Gillis et al., 2016; Goss & Hunter, 2015). There are also research articles, such that of Gorur and Wu (2015), which seek to contextualise the central intended interpretation of the PISA data for the local Australian context.

Reciprocity

Reciprocity is also evident in the PISA methodology. First are the National Involvement Standards for the PISA that require participating countries to provide feedback to ensure that the PISA test “instruments achieve cross-national, cross-cultural and cross-linguistic validity” (OECD, 2017c, p. 451). The involvement standards of the PISA also illustrate Habermas’ (1998) conceptualisation of deliberative democracy occurring through organisations and not individuals. This is illustrated by the involvement standards requiring feedback strategies to be documented and to include lists of committees and groups involved in providing the feedback, as well as membership and meeting records (OECD, 2017c, p. 451).

A second aspect of reciprocity pertaining to validation in the PISA emerges from the role of National Project Managers appointed by individual countries to implement the PISA in their jurisdiction. For the PISA 2015, these “managers played a vital role in developing and validating the international assessment instruments and ensured that PISA implementation was of high quality.” (OECD, 2017c, p. 26)

Discussion

This paper has provided a rudimentary framing for deliberative democracy in educational assessment validation through Moss’ (2016) concept of intended interpretation and actual interpretation. Having rationally reconstructed practices immanent to educational assessment, and by providing elementary verification that the identified practices are immanent in contemporary assessment programs such as the PISA, a basis is provided for critique. The method of critique emerging from rational reconstruction is generally immanent critique and critical theory (Antonio, 1981; Kellner, 1992; Stahl, 2013). These forms of critique address contradictions that emerge in practices immanent in fields such as educational assessment.

One issue that emerges from an analysis of the PISA architecture for deliberative democracy relates to the release of its test material. The PISA limits its release of test material with the majority of test material remaining secure for the purpose of trend reporting (OECD, 2017c, p. 54). That test material is only available in central contexts for validation in large-scale international assessment is considered a barrier to further analysis by Baird, Andrich, Hopfenbeck, and Stobart (2017, p. 333). The failure to release test material after it has been
administered and reported upon contrasts to traditional methods of reporting. For example, secondary school exit credentials generally make sample and past exam papers publicly available\(^2\). The failure of the PISA to release all its test material in a similar fashion erodes its legitimacy from the perspective of validation through deliberative democracy.

A further issue that emerges from trend reporting is an increasingly diminished potential for reciprocity. While the involvement standards may have been essential to ensure the initial legitimacy of the PISA material and that it had cross-national and cross-cultural validity, the potential to inform development of test material diminishes once the underlying constructs have stabilised. Trend reporting in the PISA requires the construct to remain equivalent across cycles (Dorans, Moses, & Eignor, 2011, p. 23; Gebhardt & Adams, 2007; Holland & Dorans, 2006, p. 194). For the PISA, new items need to psychometrically fit existing items to make the link between cycles psychometrically valid and stable (OECD, 2017c, p. 143). That is, the PISA construct is not able to respond to technological and social development in societies with all the PISA constructs remaining static from PISA 2000 to PISA 2015. Trend reporting results in the PISA constructs being conceptually considered as static, even when the assessment frameworks are updated for each cycle (OECD, 2000, 2003, 2006, 2010, 2013, 2017a). The effect of technological progress on student learning and student consciousness is explored by Vygotsky (1978) and Papert (1993) for example. The PISA is not able engage with these discourses of progress due to the requirement for constructs to remain equal across cycles to enable valid trend reporting.

Trend reporting results in the PISA constructs to be reified across time and diminishes the requirement for feedback on test content from countries leading to an erosion in the program’s reciprocity and legitimacy. It leads to questions such as that of Biesta (2009) who asks “whether we are indeed measuring what we value” (p. 35). Further, a diminished concern for test content is a diminished concern for the linguistic component of the PISA data resulting in an inordinate focus on its numerical component (Biesta, 2015; Gillis et al., 2016; Gorur & Wu, 2015; Grek, 2009; Lingard, 2011).

One further effect emerging from a diminished concern for test content and the linguistic component of educational assessment data is that, in Habermas’ (1992) terms, it results in “delinguistification” of the PISA data. The PISA data is best conceptualised as generalised

\(^{2}\) for example, see the Victorian Certificate of Education, Australia, www.vcaa.vic.edu.au/Pages/vee/exams/examsassessreports.aspx
communication comprising linguistic meaning reported through assessment frameworks and a numerical level indicating a country’s performance with respect to that content. When educational assessment data is decoupled from its linguistic meaning the data takes on the characteristic of abstract steering media. This results in the PISA data being used in an empirical fashion rather than an ethical fashion that addresses what is true, right, and good.

The PISA is considered by some as the most influential contemporary global assessment program (Biesta, 2015; Bonal & Tarabini, 2013; Breakspear, 2012; Grek, 2009). There is an argument to suggest that the involvement of countries through the provision of feedback demanded by the PISA technical standards might have played a part in garnering initial influence and legitimacy for the PISA. There is also an argument that the requirement for trend reporting which results in the limited release of test material and a disengagement with test content has eroded the legitimacy of the PISA over time ("OECD and PISA tests are damaging education worldwide – academics," 2014).

Conclusion

This paper has built on Moss’ (2016) conceptualisation of educational assessment validation involving an intended interpretation and actual interpretation that better addresses the use of data by professionals. The two interpretations have been conceptualised in this paper by building on Kane’s (2006) theory of validation and Habermas’ (1996, 1998) discourse ethics and deliberative democracy.

This paper has provided a framework for informing the legitimate design and use of large-scale educational assessment, particularly its reporting function. It has also provided a framework for analysing and critiquing the effect from programs such as the PISA on local contexts. The approach detailed here does not supersede the approach to validation of Kane (2006). Instead, this paper alerts to how the nature, clarity, and plausibility of arguments might differ between intended interpretation reported through central agencies and actual interpretations made by professionals in local contexts.

This paper has provided a framing for critiques of large-scale assessment that claim an excessive focus on the numerical component of large-scale educational assessment data. An inordinate emphasis on the numerical component can lead to analytic and empirical arguments emerging from the use of educational assessment data. These arguments do not address the substantial predicates in educational discourse of what is true and what is right.
argument enhances legitimacy and may require recourse to further evidence, particularly evidence drawn from local contexts.

References


