

SPECIAL ISSUE
**Learning Assessments for Sustainability? Exploring the Interaction
between Two Global Movements**

education policy analysis
archives

A peer-reviewed, independent,
open access, multilingual journal



Arizona State University

Volume 29 Number 122

September 27, 2021

ISSN 1068-2341

**Unsustainable Measures?
Assessing Global Competence in PISA 2018**

Harsha Chandir



Radhika Gorur

Deakin University
Australia

Citation: Chandir, H., & Gorur, R. (2021). Unsustainable measures? Assessing global competence in PISA 2018. *Education Policy Analysis Archives*, 29(122). <https://doi.org/10.14507/epaa.29.4716> This article is part of the special issue *Learning Assessments for Sustainability? Exploring the Interaction between Two Global Movements*, guest edited by Oren Pizmony-Levy and Dafna Gan.

Abstract: In the context of rising fundamentalism, urgent threats to the environment, and the persistence of poverty and deep inequities in the world, 193 nations have pledged to work towards the Sustainable Development Goals (SDGs) crafted by UNESCO in 2015. Education is seen as key to attaining all the other SDGs. Within the ‘education goal’ (Goal #4), there is an explicit target, SDG 4.7, which focuses on ‘sustainable development and global citizenship’. Nations are expected to incorporate a focus on SDG 4.7 into their curricula, policies, teacher education programs, and student assessment. PISA has now developed an assessment of ‘global competence,’ which is presented as a way to assess SDG 4.7. Through this assessment, it seeks to inform policy, curricula, and pedagogies and catalogue ‘best practices’ for developing students’ ‘global competence’. Given

Journal website: <http://epaa.asu.edu/ojs/>
Facebook: /EPAAA
Twitter: @epaa_aape

Manuscript received: 5/24/2019
Revisions received: 2/22/2021
Accepted: 2/23/2021

this ambition and the centrality of ‘sustainable development and global citizenship’ within the globally endorsed SDGs, it is important to analyze the extent to which the PISA assessment of global competence is usefully able to inform policy and practice and contribute to fulfilling SDG 4.7. We build upon the work of other scholars examining this question, taking a material-semiotic approach inspired by Science and Technology Studies. Empirically, our study is based on documentary analysis, interviews, and ‘survey encounters’ in which we administered a curated part of the assessment to 15-year-olds and followed this exercise with interviews. We explore how the hard-won stability gained around the notion of ‘global competence’ through its inscription into the standardized survey instruments is again threatened when the survey instrument encounters diverse 15-year-olds. The survey encounters provide an opportunity to ‘test the test’, and we conclude that the PISA test of global competence is not as yet in a position to provide useful direction to policy or practice in the promotion of SDG 4.7.

Keywords: PISA; sustainability; survey encounters; global competence; science and technology studies; international large-scale assessments

¿Medidas insostenibles? Evaluación de la competencia global en PISA 2018

Resumen: En el contexto del creciente fundamentalismo, las amenazas urgentes al medio ambiente y la persistencia de la pobreza y las desigualdades mundiales, 193 naciones se comprometieron a trabajar hacia los Sustainable Development Goals (SDGs) elaborados por la UNESCO en 2015. La educación se considera clave todos los demás ODS. Dentro del “objetivo educativo” (Goal n.º 4), el SDG 4.7 tiene como objetivo “desarrollo sostenible y ciudadanía mundial”. Se espera que las naciones incorporen un enfoque en el SDG 4.7 en sus planes de estudios, políticas, programas de formación docente y evaluación de los estudiantes. PISA desarrolló una evaluación de la “competencia global” a través del SDG 4.7 como una forma de informar las políticas, los planes de estudio y las pedagogías y catalogar las “mejores prácticas” para desarrollar la “competencia global” de los estudiantes. Este artículo analiza en qué medida esta evaluación de PISA de la competencia global es útil para informar las políticas y la práctica y contribuir al cumplimiento del SDG 4.7. Creamos una erudición interesante y adoptamos un enfoque semiótico material inspirado en los estudios de ciencia y tecnología. Empíricamente, nuestro estudio se basa en análisis documental, entrevistas y “encuentros de encuestas” en los que administramos una parte curada de la evaluación a jóvenes de 15 años y seguimos este ejercicio con entrevistas. Exploramos cómo se desestabilizó la noción de “competencia global” en este proceso. Como resultado de esta “prueba de la prueba”, llegamos a la conclusión de que la prueba PISA de competencia global no está tan preparada para proporcionar una orientación útil a las políticas o prácticas en la promoción del SDG 4.7.

Palabras clave: PISA; sustentabilidad; encuentros de encuestas; competencia global; estudios de ciencia y tecnología; evaluaciones internacionales a gran escala

Medidas insustentáveis? Avaliação da competência global no PISA 2018

Resumo: No contexto de crescente fundamentalismo, ameaças urgentes ao meio ambiente e a persistência da pobreza global e das desigualdades, 193 nações se comprometeram a trabalhar para os Sustainable Development Goals (SDG) elaborados pela UNESCO em 2015. A educação é vista como a chave para atingir todos os outros ODS. Dentro da “meta de educação” (Goal 4), o SDG 4.7 visa “desenvolvimento sustentável e cidadania global”. Espera-se que as nações incorporem um foco no ODS 4.7 em seus currículos,

políticas, programas de formação de professores e avaliação de alunos. O PISA desenvolveu uma avaliação de ‘competência global’ por meio do SDG 4.7 como uma forma de informar políticas, currículos e pedagogias e catalogar as ‘melhores práticas’ para desenvolver a ‘competência global’ dos alunos. Este artigo analisa até que ponto esta avaliação do PISA de competência global é útil para informar políticas e práticas e contribuir para o cumprimento do SDG 4.7. Construimos bolsas de estudo interessantes e adotamos uma abordagem semiótica material inspirada nos Estudos de Ciência e Tecnologia. Empiricamente, nosso estudo é baseado em análise documental, entrevistas e “encontros de pesquisa”, nos quais administramos uma parte da avaliação com curadoria de jovens de 15 anos e seguimos esse exercício com entrevistas. Exploramos como a noção de “competência global” foi desestabilizada neste processo. Como resultado deste ‘teste do teste’, concluímos que o teste PISA de competência global não está tão pronto para fornecer uma direção útil para a política ou prática na promoção do SDG 4.7.

Palavras-chave: PISA; sustentabilidade; reuniões de pesquisa; competição global; estudos de ciência e tecnologia; avaliações internacionais em grande escala

Unsustainable Measures? Assessing Global Competence in PISA 2018

When 193 nations of the world came together to become signatories to UNESCO’s Sustainable Development Goals (SDGs), they committed to achieving a series of targets that together spearheaded a global ‘sustainable development movement’. The 17 SDGs include a wide range of goals, such as ending poverty, achieving gender equality, achieving sustainability in water management and sanitation, and taking urgent action to combat climate change. The global community agreed to reach these goals by 2030, and to measure and monitor their progress towards these goals to ensure they achieved the targets they had set. Seeing education as the key to achieving all of the 17 SDGs, UNESCO has incorporated the sustainability agenda within the education goal (SDG4), in the form of Target 4.7, which reads:

By 2030, ensure that all learners acquire the knowledge and skills needed to promote sustainable development, including, among others, through education for sustainable development and sustainable lifestyles, human rights, gender equality, promotion of a culture of peace and non-violence, global citizenship and appreciation of cultural diversity and of culture’s contribution to sustainable development. (UNESCO, 2016, p. 13)

The SDGs are also keenly focused on monitoring the progress towards achieving these goals. To aid such monitoring, UNESCO has developed a ‘global indicator’, SDG 4.7.1, which reads:

Extent to which (i) global citizenship education and (ii) education for sustainable development, including gender equality and human rights, are mainstreamed at all levels in: (a) national education policies, (b) curricula, (c) teacher education and (d) student assessment. (Sandoval-Hernández et al, 2019, p. 3)

Five ‘thematic indicators’ follow this, of which we focus here on two:

4.7.4 Percentage of students by age group (or education level) showing adequate understanding of issues relating to global citizenship and sustainability.

4.7.5 Percentage of 15-year-old students showing proficiency in knowledge of environmental science and geoscience. (Sandoval-Hernández et al., 2019, p. 3)

These two indicators were linked to international large-scale assessments, according to the UNESCO proposal for measuring SDG4.7:

In fact, it is important to notice that, according to the measurement strategy for SDG target 4.7 proposed by GAML [Global Alliance for Monitoring and Learning] and the Task Force 4.7 (UIS, 2017), the two thematic indicators were originally inspired by the IEA's International Civic and Citizenship Study (ICCS) and the OECD's Programme for International Student Assessment (PISA). (Sandoval-Hernández et al., 2019, p. 3)

UNESCO had been interested in the notion of global citizenship education, which can be seen as one of the precursors to SDG 4.7, for decades. In 2013, UNESCO commissioned a paper titled *Measurement of Global Citizenship Education* (Skirbekk et al., 2013), and while it cited a number of existing scales and made recommendations for measuring global citizenship education, it made no reference to PISA and contained only a fleeting reference to ICCS (and then only to note that it was inadequate as it only covered a small group of countries). However, it appears that at some stage ILSAs had begun to influence the formation and elaboration of SDG 4.7, as evidenced in the document cited above.

This relationship was not one way. For its part, the OECD, which originally had a strongly economic focus in its notion of 'global competence' (GC) when it began to consider, in 2013, adding it to the 2018 assessment as the cross-curricular dimension, now explicitly aligns itself with the UNESCO 2030 program and the sustainability agenda (see Auld and Morris 2019 for a detailed analysis). Introducing the PISA measure of GC, OECD's Director of Education and Skills, Andreas Schleicher, endorses the strong emphasis placed in SDG 4 on 'learning to live together sustainably'. He adds:

But such goals are only meaningful if they become visible. This has inspired the OECD Programme for International Student Assessment (PISA), the global yardstick for educational success, to include global competence in its metrics for quality, equity and effectiveness in education. PISA will assess global competence for the first time ever in 2018. In that regard, this framework provides its conceptual underpinning. (OECD, 2018, p. 2)

The PISA assessment of GC, then, represents a nexus between the sustainability movement and the ILSA movement. Through the PISA survey of GC, the OECD claims to make visible the extent to which SDG 4.7 is being achieved in a globally comparative way. It also declares that:

A fundamental goal of this work is to support evidence-based decisions on how to improve curricula, teaching, assessments and schools' responses to cultural diversity in order to prepare young people to become global citizens. (OECD, 2018, p. 6)

In this paper, we examine this claim and ask: What notions of sustainability are inscribed into the survey instruments? And to what extent are these ideas meaningfully translated into survey questions that would yield useful data? Based on this analysis, we seek to answer the question: does the PISA survey of global competency offer useful information that could help nations make appropriate policy changes to educate students for a more sustainable world?

Our analysis is in two main parts. The first part explores how the slippery notion of ‘global competency’ is stabilized through its translation into a set of survey instruments. We follow how the OECD team charged with developing this survey of global competencies negotiated a series of political and epistemological challenges to eventually produce instruments that held together, however tentatively, a collectively endorsed idea of ‘global competency’. The second part of our analysis involves an examination of what happens when this particular set of understandings of GC inscribed into the survey instruments encounter the 15-year-olds taking the survey. Using a ‘sociology of quantification’ approach inspired by Science and Technology Studies (STS) in our analysis, we argue that the concept of ‘sustainability’ envisioned in SDG 4.7 is not easily translated into universally measurable ‘global competencies’.

In the next section, we elaborate our methodological stance, linking it to the work of other scholars focusing on the PISA GC assessment. Two analysis sections follow, the first concerned with the production of the GC surveys and the second with the encounter between the laboratory production of GC and 15-year-old students from three Melbourne schools. We conclude by summing up our findings and engage in a discussion of the larger issues connected with quantification, and the inherent tension between the broad focus on inclusivity and the valuing of diversity that is the thrust of the SDGs, and the standardization, harmonization and alignment that is integral to both the ILSA movement and to the work of global development agencies.

Theoretical Framework and Methodology

We draw on the rigorous work of a small group of scholars who have been writing about the PISA GC assessments. Auld and Morris (2019) argue that in developing the assessment of GC and aligning it with the UNESCO 2030 agenda, the OECD makes a strategic and political move to further its capacity for epistemic governance (Sellar & Lingard, 2014) as PISA expands to more and more nations, including those in the global south. Analyzing the changing definitions of ‘global competence’ within the OECD between 2013 and 2018, they trace how the initial focus on how well-equipped students are to thrive in a global labor market, proposed in the 2014 framework, is transformed – but more in words than in spirit – to one that is more closely aligned with the SDG conceptualization of global citizenship.

Based on *Global Competency for an Inclusive World* (OECD, 2016), a document that has since been taken off the OECD website and replaced by *Preparing Our Youth For An Inclusive And Sustainable World* (OECD, 2018), Ledger et al. (2019) note that the philosophical underpinnings of the OECD’s take on GC is rather narrow, and that the kind of 15-year-old envisaged in the GC assessments is one who is privileged, globally mobile, and able to exert agency on the world. They argue that students from low-income backgrounds attending schools that do not offer additional languages or overseas student exchange programs would be seen as lacking in GC on this measure. They also point to the small range of people whose ideas have influenced the concept of GC, based on a social network analysis.

While Auld and Morris and Ledger et al take a political stance, focusing on the motivations and geopolitical ambitions of the OECD and critiquing the thinly disguised economic agenda beneath the links made to the humanitarian and sustainability agenda of the SDGs, Sälzer and Roczen (2018), who were both involved in developing the PISA measures of GC, provide a critique based on a technical analysis. Outlining what it takes to develop a technically sound set of measures, they point to the impoverished theoretical development of the concept of GC, and the consequences that flow from this lacuna. They suggest that it is both important and possible to develop good measures of GC, but that the current measure falls far short of ideal.

Locating our study in the interdisciplinary field of Science and Technology Studies (STS), which is interested in knowledge practices in all its forms (Jasanoff, 2004), we straddle both the ‘political’ and the ‘technical’ aspects of measurement – in other words, we view such assessments as socio-technical (Gorur, 2017). We trace the processes by which the scientific feat of establishing an international, comparative measure of GC is achieved. We follow the efforts to define and stabilize a nebulous, uneven concept and elaborate it into a series of indicators that can be measured. In this, we follow the lead of a range of historians and sociologists of science who have traced how facts and numbers are produced (for example, Desrosières, 1998; Hacking, 1990; Poovey, 1998; Porter, 1996). Our objective is to perform a sociology of quantification (Derksen, 2000; Espeland & Sauder, 2007; Gorur, 2014; Hardy, 2015; Woolgar, 1991) that traces how certain concepts come to be defined, elaborated into observable and measurable indicators, and translated into trusted measures that eventually come to be seen as objective and useful and begin to confidently circulate through scientific communities and societies at large (or, in some instances, fail to do so).

In his studies of laboratory science, Latour (1987) has made two key points that underpin our approach. First, he asserts that ‘finished science’ (we use ‘science’ here to mean ‘knowledge’) – the ‘facts’ that scientists produce, such as ‘the structure of the DNA is a double-helix’ – are not just a matter of ‘discovering’ a correspondence with nature, but a more or less fragile assemblage that is a result of funding, scientific protocols, laboratory instruments, the cooperation of colleagues, etc. Here, he sees science as Janus-faced – post facto, we could argue that we agree with Crick and Watson¹ when they say the structure of DNA is a double helix because this idea corresponds with nature. But tracing scientific practice historically, we could say that the structure of the DNA is a double helix because Crick and Watson were able to assemble a string of actors to gain agreement that this is the case.

Similarly, we take nothing for granted about the development of valid protocols to measure GC in an internationally comparable way. We do not take ‘GC’ to be something prior to its definition and measurement so that we can look at the final product and see if it can faithfully measure GC. Rather, we posit that the act of defining and measuring GC *performs* it into being, stabilizing it as a particular set of ideas. Because this exercise involves performing a particular notion of GC rather than another, this is a *productive*, rather than representative exercise (Knorr-Cetina, 1981). In other words, this involves what Annemarie Mol (1999) calls ontological politics. It is this productive aspect that makes such activities worthy of investigation. If a different definition or a different set of indicators had been adopted, the results of the GC measures might have worked out very differently. The role of instruments in generating findings has been well documented. For example, Gorur (2014) has shown that a small change in the way ‘equity’ was measured in PISA meant that Australia went from being classified as ‘low equity’ in PISA 2000 to ‘high equity’ in PISA 2003.

In this approach, changing definitions until the team has arrived at one that gains enough consensus to get approval and funding to proceed with the measurement is not scandalous in itself; as many studies of laboratory science have shown, and as, indeed, Sälzer and Roczen (2018) have also shown, this is how ‘science’ (i.e., knowledge making) is done. Peer review epitomizes this requirement of gaining the agreement of fellow experts for an idea to stabilize and become accepted. It is elaborating this ‘constructed’ or ‘fabricated’ nature of science that has been a key contribution of STS. What is important, then, is to explore how ‘matters of fact’ come to be made, and to understand the implications and ramifications of the particular construction under study on the

¹ We would like to acknowledge the work of Rosalind Franklin, whose essential contributions to the discovery of the structure of DNA are often overlooked.

different groups and actors it affects, and on the knowledge that is produced. To understand the process by which the concept of GC came to be defined and stabilized in PISA, semi-structured interviews were conducted via Skype by Author #1 (Chandir) with ‘elite’ informants (Addey, 2019), who included key OECD officials and members of PISA 2018’s Global Competence Expert Group (GEG) who agreed to be interviewed.

The story of the stabilization of the notion of GC through its inscription into a set of questionnaires is, however, just part of the story. The fate of objects – including a particular rendering of GC – remains in the hands of its users. It is only when the carefully assembled human and non-human actors circulate through the world, encountering a range of other actors and facing the trials they impose, that we can gauge the strength of that assemblage (Latour, 1987). As the GC assemblage, now gathered in a particular way in the questionnaires, makes its way through the world, it meets new actors who may challenge or transform it in different ways. The ‘objectivity’ of science, according to Latour (1987), is gained by this process, in which ‘objections’ are able to be raised. Sometimes the object has to change – or be translated – in the face of such trials. At other times, it manages to convince those it encounters of its soundness. In the case of the PISA GC assessment, we were curious to see what happened when the questionnaire encountered its most crucial audience – the 15-year-olds for whom it was designed.

Serder and Jakobsson (2015) studied students as they worked in groups of three to answer a selection of cognitive science items from PISA 2006. They refer to this as ‘a study of student-assignment encounters in action’ (p. 833). The purpose of getting the students to respond in groups was to make observable how the students ‘negotiate[d] the meaning of the questions’ (Serder & Jakobsson 2015, p. 865) and made some of their ‘thinking processes more explicit’ (Le Hebel et al., 2014, p. 278), which cannot be captured in a written test where individual students quietly fill out the answers.

Building on the above studies, Author #1 developed the method of ‘survey encounters’ (Chandir, 2020). Using a curated set of the open-access items in the GC attitudinal survey, the questionnaire was administered to PISA-age students (students between 15 years, 3 months and 16 years, 2 months) in groups of three. The students were requested to collaboratively, and through discussion, fill out the responses to the questions presented. The sample included a government school (Springfield)², an independent school (Chilton) and a Catholic school (Malory).³ The students’ discussions as they read and attempted to make sense of the questions were audio recorded. Later, the students were interviewed to probe in greater depth some of the issues raised in the group ‘survey encounters.’ Interviews were audio recorded, and data analyzed by identifying common codes across the participants’ responses from the survey encounters and semi-structured interviews.

The three papers already published on the PISA GC assessment – Auld and Morris (2019), Ledger et al. (2019), and Sälzer and Roczen (2018) – were all written before the PISA GC assessment was finalized and administered. We follow the story a bit further – at the time of this writing, the PISA assessment of GC has already been conducted, and the self-reported non-cognitive survey is available to download online, as are some sample scenarios and questions from the cognitive questionnaire. These new materials, our theoretical framework, and the inclusion of the empirical materials – the survey encounters and the interviews with the 15-year-old test takers and

² All schools have been given pseudonyms.

³ The purpose of choosing different types of schools was to explore the effects of different curricula and on school-related factors on GC education, and to investigate this, Author #1 also administered the teacher and principal survey. That part of the study is beyond the scope of this paper.

with the designers of the assessment – allow us to make original contributions to the accumulating work on the PISA GC assessments.

“Global Competency”: From a Nebulous Idea to an Internationally Comparable Survey

Before something can be measured, it needs to be defined to ensure that the measurement instruments target the right concept. Constructs come to gain definition through theoretical work, which involves trials in the form of peer review and epistemic debates. Even now-established concepts such as ‘literacy’ and ‘numeracy’ once were elusive and challenging to stabilize across cultures and languages. Moreover, definitions do not just stay in place once they are established – there is a need for on-going labor to maintain or update these definitions with new developments in literacy and numeracy theories and practices (cf. Sauvageot, 2008).

In the case of ‘global competence’, there is little agreement on the boundaries of the concept, let alone definitions or underpinning theories. Debates about the purposes of education are ever present in every society, and they usually include some variation of readying students for the workplaces they will encounter and inculcating the kinds of values and behaviors that will make them good citizens and dispose them towards acting for a peaceful and just world. A range of notions such as global citizenship, multiculturalism, 21st century skills and international mindedness have circulated in various influential policy documents (see, for example, Deardorff, 2006; International Baccalaureate Organisation, 2015; OECD, 2018; UNESCO, 2014; Voogt & Roblin, 2012). Which aspects are recognized and what weightage is assigned to the different factors that make up the idea of ‘global competence’ depends very much on the institutions and the individuals defining the concept and the contexts within which they operate.

‘Definition’ has a precise ring to it – it suggests that arriving at the ‘right definition’ – one that precisely describes a concept – is a matter of wordsmithing until a description that most closely matches the phenomenon is found through the consensus of relevant groups. But what if the concept itself is fuzzy and nebulous? How can experts ‘arrive’ at a definition, as if the object already existed prior to its definition? Following Latour’s notion of the Janus-faced nature of scientific work, we would suggest that definitions, if they hold, beget the things defined.

In PISA, the definitional work is done through the development of an ‘assessment framework’ for the domain in question. The PISA domain frameworks go beyond definitions – they are ‘constructs’ which are more elaborate than a mere definition and serve to outline how the experts envision the operationalization of the concept:

Every assessment framework used in the context of PISA is developed along three criteria:

- (1) the construct of interest needs to be relevant for everyday situations;
- (2) it needs to represent a central domain-specific theory or a basic concept of the domain; and
- (3) it needs to be appropriate for capturing the developmental stage of 15-year-olds (OECD, 2016⁴, cited in Sälzer & Roczen, 2018).

The framework for GC, then, needed to not only conceptualize a particular understanding of GC for which it could get consensus among the experts and the approval of the PISA Governing Board,

⁴ PISA 2015 *Assessment and Analytical Framework: Science, reading, mathematics and financial literacy*.

but also ensure that this construct was linked to sound theory and had relevance to the everyday situations of 15-year-olds.

The crafting of this framework is a critical step because it shapes and stabilizes a particular construct of the concept. Once the domain Expert Group – in this case the Global Competence Expert Group (GEG) – develops the framework, it is used as a screening device against which the ‘fit’ of the test items is determined. It thus serves to consolidate the favored construct by generating the cognitive test items and the contextual items that align with it, and keeps out items that do not fit and which might destabilize the construct if admitted into the assessment (Gorur, 2011).

In the rest of this section, we trace the ways in which the construct and framework of GC came to be elaborated and contested until a particular iteration came to be inscribed into the 2018 instruments.

Global Competence, the OECD and PISA

In 2013, Fernando Reimers, an expert who contributed to UNESCO’s (2013) *Technical Consultation on Global Citizenship Education*, presented a paper to the OECD which outlined the importance of developing an assessment of global competence, and presented it as an opportunity for PISA (see also Auld and Morris, 2019). In this paper, titled *Assessing Global Education: An Opportunity for the OECD* (Reimers, 2013), he suggested that as individuals and nations are becoming more interdependent, interconnected and increasingly engaged in their interactions with others from different cultures, it is imperative to ‘figure out their common humanity and their differences with others’. The ‘new civics’ envisaged by Reimers included education for human rights, intercultural understanding, and development studies.

It is not clear whether Reimers’ paper had a significant influence on the OECD, but the idea of measuring ‘global competence’ was already underway around this time, with a view to developing an assessment of GC as a ‘cross curricular domain’ to add to its suite of assessments in 2018 (Sälzer & Roczen, 2018). OECD had appointed Pearson, who were already contracted to develop various other aspects of the PISA survey, to do some of the background work for assessing GC. Reimer’s paper was referenced, among many others, in the *PISA 2018 Framework and Plans* document prepared for the 38th meeting of the PISA Governing Board (PGB) meeting in October 2014.

The construct of GC in this 2014 document presented to the PGB differed sharply from the ways in which UNESCO (and Reimers) had been conceptualizing this idea. Rather than focusing on how individuals might work to serve a common humanity, reduce inequities and foster a just world in which diversity is valued, the 2014 *Framework and Plans* document conceptualized GC as the knowledge, skills and attitudes that students needed to thrive in a world where they would likely be studying and working overseas. In other words, rather than individuals in the service of humanity, the focus was on how individuals could get maximum value and benefit from a globalizing world.

This Framework argued that students needed an education in which they were exposed to ‘international and intercultural partnerships’, rationalizing this need as follows:

Such experiences are crucial for the growing number of students who will seek to pursue their further studies outside their home country. Moreover, when they enter employment they will be confronted with the international nature of the world of work and the global way in which trade, finances and laws increasingly operate. (OECD, 2014, p. 6)

Clearly, as Ledger et al. (2019) point out, the student Pearson was imagining was a privileged one. This student is destined to take advantage of the ‘increased travel opportunities, increased student mobility and the proliferation of communication devices’ and able to contribute to the 7 trillion-

dollar travel industry and be among the 4.5 million tertiary students studying overseas (OECD, 2014, p. 6).

Assembling the Construct

The Framework and Plans document explains how Pearson went about this task: Because GC is an important and emerging domain appearing for the first time in PISA 2018, Pearson engaged in the following focused work on this domain:

- Between December 2013 and February 2014, Pearson's Research and Innovation Network conducted a literature review and prepared a white paper on the area of global competence
- Between January and March 2014, Pearson held several virtual meetings and written exchanges with international experts to discuss the scope, direction and potential issues related to the new global competence domain... (OECD, 2014, p. 7)

The experts involved are listed in the framework document. Linking with established science and listing the names of the experts is important in stabilizing a construct, gaining credibility and battling away sceptics. At the same time, demonstrating the 'workability' of the model is also crucial – particularly to gain the support of the PGB, without which the assessment cannot proceed. Thus, this work is grounded in research but also in the pragmatics of what works in PISA:

The development of the framework builds on this earlier work and is coordinated with the work of the reading literacy expert group (REG) and questionnaire expert group (QEG). This work continues to draw heavily on Pearson's existing expertise and knowledge of what works in PISA, in addition to the input of a range of experts who have knowledge of different dimensions of global competence. (OECD, 2014, p. 8)

A detailed outline of Pearson's work in assembling the expertise to create a robust model of GC upon which an assessment framework could be constructed is outlined for the benefit of the PGB. The working definition of GC offered in this document to the PGB read as follows:

Global Competence is the capacity of an individual to understand that we learn, work and live in an international, interconnected and interdependent society and the capability to use that knowledge to inform one's dispositions, behaviors and actions when navigating, interacting, communicating and participating in a variety of roles and international contexts as a reflective individual. (OECD, 2014, p. 9)

Absent in this definition is any hint of the vision of SDG 4.7, in which the individual is working for the welfare of all, interested in reducing inequalities and focused on the health of the planet, promoting sustainability and so on (see also Auld and Morris, 2019).

The Cultural Perspectives of Middle-Aged Americans?

Pearson had taken great pains to make visible the scientific basis of its work, declaring a long list of experts involved, referencing a range of literature, linking with existing theory including validity theory, and preparing complex diagrams, with passages such as:

Generally, the new global competence framework is being developed by following the research-based development matrix summarized in Figure 2. This development matrix is grounded in validity research (e.g., AERA, APA, & NCME, 1999; Kane, 1992; Messick, 1989; Mislevy, Almond, & Lucas, 2003) and represents a road map

for gathering and organizing an evidentiary basis to underpin the validity of the definition of a domain and its operational measurement. (OECD, 2014, p. 10)

Despite all this work to present the construct as technically and scientifically sound, according to Sälzer and Roczen (2018) and also the PISA GC expert interviewee, this construct did not go down well with the PGB, and many countries decided to not participate in the assessment of GC after all. The PISA GC expert suggested that part of the issue with the first GEG was its lack of diversity. It was made up of experts from England, Finland, South Africa, USA and Germany. In describing this group of experts, Anke Grotlüschen states:

Overall, the group never was very global; it never included countries which are less familiar with the English language than the former Commonwealth and Asian countries. There is no contribution from Latin America and Africa except Sarah Howie, but she is in charge of linking the domain with her expert group on the background questionnaire. (Grotlüschen, 2018, pp. 195-196)

Grotlüschen argues that though many of these experts had worked in the field of global citizenship, global competence and intercultural competence, the dominant discourse seemed to be ‘western, cognitive and rational’ (Grotlüschen, 2018, p. 196). For example, empathy is positioned as a skill to have rather than something one feels.

An expert interviewee provided an understanding of the kind of issues that came up. A set of questions were based on a chat format, where students were expected to state if they would be willing to engage in online chats about topics such as the construction of dance in South India or fair wages for workers. The interviewee suggested that the GEG had been unimaginative in their assumption that ‘15-year-olds all around the globe engage voluntarily in a school break [in a] discussion, with their peers, about issues like pineapples and fair wages for workers or dance in Southern India’ (interview transcript, PISA GC expert). Not only the content, but the language used was also not authentic – in part because they (unlike 15-year-olds) had to use language that was both grammatically correct and appropriate. The cultural distance between the ‘middle-aged American test developers’ and the intended test takers posed a major challenge:

We are already asking them to play a game they probably didn’t play or didn’t find quite familiar... and then we are asking them to say something and to use words that basically [were] predefined with a certain cultural perspective, [that] of the test developers... [the] middle aged, American test developers. (Interview transcript, PISA GC expert)

When the PGB rejected the Pearson construct and Framework, the design and development of the framework was taken over by an in-house PISA analyst who brought together a new group of experts. However, in order to keep the continuity of the process, one of the members of the first GEG was invited to be part of the second GEG.

These events epitomize the kind of trials to which attempts to stabilize an idea are subjected. Pearson’s efforts to legitimize its construct as ‘science’ – as trustworthy knowledge – did not succeed. The construct did not gain the status of ‘objectivity’ since it could not withstand the ‘objections’ raised by the PGB. One GEG member said of the two groups:

The first group was bigger, it was international, from different countries. Not all of the persons were what I’d consider ‘the experts’ in Intercultural or Global Competence. I don’t know how they were appointed. I don’t know how they were brought together... [The second group] was much smaller and we all already knew

each other. We already had established relationships with each other, so it was probably much easier to work with each other... and in the end it was four of us representing the US, the UK, Argentina in some ways, Korea and [a PISA analyst]. (Interview Transcript, GEG member)

There is an important point to note here. Whereas after-the-fact ('cold science') accounts simply present a framework or a calculation as a matter of rational, scientific thinking, in-the-moment ('hot science') accounts are full of 'feelings' as well as 'science'. Research by Gorur has shown time and again that when teams – like the GEG – are formed, choices of whom to include are often made on the basis of having worked together on past projects or having attended the same university. The choices appear to be based both on belonging to the same epistemic community and knowing that the members will likely get on with each other. When working for hours struggling over precise semantics to gain consensus from the whole group, it helps if the group is small, and if the members of the group like and trust each other, and perhaps socialize after work, especially since these meetings are often held in some foreign city. Having this comfort and level of personal friendship is important in reaching consensus and stabilization (Gorur, 2015; Sauvageot, 2008). It is within these smaller and more homogeneous 'epistemic communities' (Haas, 1992; Knorr-Cetina, 1999) that particular understandings and practices get stabilized and perpetuated, whereas a more diverse group might struggle to do so. However, this smaller group would indicate a less diverse view of what GC encompasses. This concern, raised also by Auld and Morris (2019), is succinctly summed up by Ledger et al. (2019):

...despite OECD nominally espousing a version of global competency based on multiple perspectives and understanding cultural differences, [our] findings show evidence of an OECD conversation impoverished by a limited degree of diversity of scholars, publication types, backgrounds, and viewpoints (p. 27).

This, they point out, is 'a mistake that runs afoul of the very purpose of educating for global competency (Ledger et al., p. 24). Ironically, in identifying 'global' competencies, the framework developed by the first GEG (described as being 'bigger' and 'international') was not well received by the representatives of 80 countries. However, the one developed by the second smaller and more homogenous group was the one that eventually was accepted by the PGB.

Take #2: GC in 2016

The work of the second expert group resulted in the 2016 OECD report *Global Competency for an Inclusive World*, and it framed GC as follows:

Global competence is the capacity to analyze global and intercultural issues critically and from multiple perspectives, to understand how differences affect perceptions, judgments, and ideas of self and others, and to engage in open, appropriate and effective interactions with others from different backgrounds on the basis of a shared respect for human dignity. (OECD, 2016, p. 4)

In this framing, there is a nod to diversity in the form of 'multiple perspectives' and 'differences', as well as the inclusion of 'shared respect and human dignity'. Still absent in this framing is the UNESCO concern with sustainability and an active concern for the wellbeing of humankind. Despite this, Schleicher was already linking the construct with the SDGs, writing that

the OECD views GC as being shaped by the three values of equity, cohesion and sustainability. Under sustainability, Schleicher states:

Delivering on the UN Sustainable Development Goals is a priority in the international community. Development that meets the needs of the present, without compromising the ability of future generations to meet their own needs, is more relevant than ever before, in the face of environmental degradation, climate change, over consumption and population growth. (Schleicher, 2016)

This report has since been taken off the Internet and in its place is a new proposal *Preparing our Youth for an Inclusive and Sustainable World* (OECD, 2018). However, the paper by Sälzer and Roczen provides a thorough and useful analysis of the 2016 framework, explaining that at that stage the theoretical underpinning of the concept of ‘global competence’ was undercooked, and as a result, operationalizing it into a rigorous set of assessment tools was difficult.

The 2018 Iteration: Alignment with SDG 4.7

By 2018, the construct of GC came to be elaborated extensively. In *Preparing our Youth for an Inclusive and Sustainable World*, the OECD lays out an impressively confident and detailed framework that includes a three-page introduction. It emphasizes the importance of having an international assessment of GC, details the concept, elaborates its dimensions, lays out the ‘building blocks’ of ‘knowledge, skills, attitudes and values,’ and provides an explanation of the assessment strategy. While the cognitive assessment items included in the actual survey are not disclosed, ‘illustrative examples’ of the scenarios are provided. References run into four pages, and the entire document, available online, is 62 pages long. The definition that guided the assessment of GC, as stated in the 2018 brochure is:

Global competence is the capacity to examine local, global and intercultural issues, to understand and appreciate the perspectives and world views of others, to engage in open, appropriate and effective interactions with people from different cultures, and to act for collective well-being and sustainable development. (OECD, 2018, p. 7)

The notion of acting for collective wellbeing is now included in this definition, as is the term ‘sustainable development’. There is an explicit reference to the SDGs in the foreword by Andreas Schleicher, and ‘supporting SDG4 (with a specific reference to 4.7) is stated as one of the goals of the assessment of GC.

Auld and Morris (2019) note that despite the effort to link with the humanitarian and egalitarian aspects of GC, the PISA conceptualization remains, at heart, committed to an economic, instrumentalist approach. We get glimpses of this focus on individual gain every now and again in the document. For example, the opening paragraph of the introduction reads:

Twenty-first century students live in an interconnected, diverse and rapidly changing world. Emerging economic, digital, cultural, demographic and environmental forces are shaping young people’s lives around the planet, and increasing their intercultural encounters on a daily basis. This complex environment presents an opportunity and a challenge. Young people today must not only learn to participate in a more interconnected world but also appreciate and benefit from cultural differences. (OECD, 2018, p. 4)

Here, ‘opportunity’ and ‘challenge’ are tied to learning to ‘participate in an interconnected world’ and to ‘appreciate and benefit from cultural differences.’ The ‘gain’ here is individualistic –

how individuals can manage to thrive and maximize individual benefit. The focus on ‘collective well-being’ and ‘sustainability’ is absent in this introductory paragraph, although it features in many other parts of the document, for example in defining the ‘globally competent student’ as one who:

can examine local, global and intercultural issues, understand and appreciate different perspectives and world views, interact successfully and respectfully with others, and take responsible action toward sustainability and collective well-being. (OECD, 2018, p. 4)

Four reasons are presented under ‘Why do we need global competence?’:

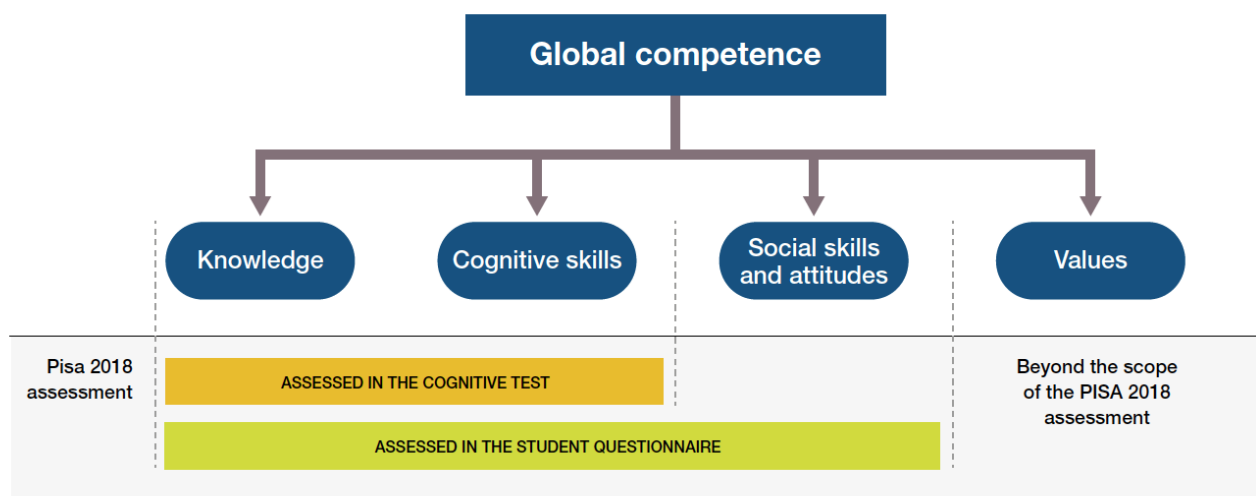
- To live harmoniously in multicultural societies
- To thrive in a changing labour market
- To use media platforms effectively and responsibly
- To support the Sustainable Development Goals. (OECD, 2018, pp. 4-5)

The first three dot points carry hints of the first iteration, with its focus on thriving in international labor markets by acquiring communication skills and being able to navigate through a diverse and multicultural world. But the last dot point is now presented as if it existed from the start as one of the rationales for developing the PISA assessment of GC.

Operationalizing the framework involves elaborating the definition so that it is relevant to everyday life and meaningful to 15-year-olds. Notions such as having the capacity to examine a wide range of issues, understanding and appreciating different perspectives, interacting appropriately with people from different cultures and acting in sustainable and responsible ways now need to be linked to the knowledge, skills, attitudes and values deemed to be required to do those things. Then items have to be devised to elicit responses that can be evaluated and graded in meaningful ways. Figure 1 provides a schematic picture of the assessment:

Figure 1

The PISA Approach to Assessing Global Competence



Note. Reprinted from OECD, 2018, p. 22.

As Figure 1 explains, while values were considered important, they were beyond the scope of the 2018 assessment. The aims of the cognitive survey and the student questionnaire are elaborated as follows:

The cognitive assessment is designed to elicit students' capacities to critically examine global issues; recognise outside influences on perspectives and world views; understand how to communicate with others in intercultural contexts; and identify and compare different courses of action to address global and intercultural issues... In the background questionnaire, students will be asked to report how familiar they are with global issues; how developed their linguistic and communication skills are; to what extent they hold certain attitudes, such as respect for people from different cultural backgrounds; and what opportunities they have at school to develop global competence. (OECD, 2018, p. 6)

The PISA 2018 brochure describes the four aspects of GC as 'supported by four inseparable factors: knowledge, skills, attitudes and values' (OECD 2018, p. 12). An explanation is provided to show the interconnectedness of the factors:

For example, examining a global issue (dimension 1) requires knowledge of a particular issue, the skills to transform this awareness into a deeper understanding, and the attitudes and values to reflect on the issue from multiple cultural perspectives, keeping in mind the interest of all parties involved. (OECD 2018, p. 12)

We would pause here briefly to reflect on the work done by these brochures. The publication of such brochures where concepts are clearly outlined, processes explained, references provided, and examples given to enhance understanding are also trust-building mechanisms. Such documents assemble a motley of actors – established theory, international experts, past experiments, diagrams and models – to prove the thoroughness of their approach. The more connections they demonstrate, and the more expertise they assemble, the greater the trust they are likely to generate.

A footnote elaborating that the four factors were based on the work of the Council of Europe exemplifies this shoring up of credibility through links with other experts and the detailing of process:

The discussion regarding knowledge, attitudes, skills and values in this section draws upon the conceptualisation of these components provided by the Council of Europe (2016) which was developed through an extensive process. It involved auditing 101 existing conceptual global, intercultural and civic competence schemes. The basic values, attitudes, skills, knowledge and understanding throughout the schemes were then identified, and a set of criteria identifying the core values, attitudes, skills, knowledge and understanding was established. Next, a first draft of the resulting model was produced and academic experts, education practitioners and policy makers reviewed and endorsed the model. It was then fine-tuned and finalised, taking into account the experts' feedback.

Noteworthy is the reference to the auditing of 101 existing schemes similar to GC. Paradoxically, providing cautions about limitations can also encourage trust; it reassures readers that the promoters are not making tall claims that cannot be supported, and this generates trust in their ethical disposition. The 2018 brochure, for example, staves off potential criticism by saying that it does not

pretend the description of the knowledge, skills attitudes and values it outlines are ‘conclusive’ or ‘omni-comprehensive’ (OECD, 2018, p. 12).

The Gradual, Partial, Stabilization of GC

The shifting rationale of the OECD in developing GC can be seen from its different articulations over the years: Reimers’ ‘new civics’ emphasizing educating for human rights, Pearson’s focus on human capital development, the 2016 proposal’s focus on respectful relationships, and the inclusion of the SDGs in the 2018 definition. The confident publication of 2018 (‘finished science’) provides no hint of the uncertainties that attended the construction of the definitions and the framework of this concept (‘hot science’). Disagreements and messy politics are swept under the carpet. Although the alignment with the SDGs occurred at the end of the process, it is presented as one of the goals of the assessment. In the preface, Schleicher makes an explicit link to the SDGs, saying

In 2015, 193 countries committed to achieving the UN’s 17 Sustainable Development Goals (SDGs), a shared vision of humanity that provides the missing piece of the globalisation puzzle. The extent to which that vision becomes a reality will depend on today’s classrooms... This has inspired the OECD’s PISA, the global yardstick for educational success, to include global competence in its metrics for quality, equity and effectiveness in education. PISA will assess global competence for the first time ever in 2018. In that regard, this framework provides its conceptual underpinning. (OECD, 2018, p. 2)

The shifting views and assumptions about what it means to be globally competent show how difficult it is to stabilize such concepts, and what a lot of effort is required to assemble the complex network of heterogeneous elements that are attempting to shape and stabilize it. Despite all this labor, the international community has not proven easy to convince. When the assessment was finally implemented in 2018, only 27 nations took both the cognitive and non-cognitive tests, with a further 29 taking the non-cognitive part only, of the total of 80 who participated in PISA 2018. Fifteen of the 80 countries declined to participate in either⁵.

In an interview with a senior member of the OECD, Author #1 (Chandir) asked if there was some anxiety or disappointment with the fact that nations had not embraced the assessment wholeheartedly. Not all, the interviewee replied, pointing out that when PISA first started its surveys of the three literacies, there were very few countries interested in participating, whereas it has now grown to include 80 nations in the 2018 survey, with plans for continued expansion.

The stability of such practices of quantification are thus a matter of time and habit and the development of confidence. The PGB members would no doubt exchange notes about whether the assessment of GC worked well and was worthwhile. Various considerations including political expediency and cost would also play a part in determining whether nations would participate in the next round of this survey.

In this section, we explored how the fuzzy and nebulous concept of GC came to be gradually stabilized – first through a definition, then through an elaboration of that definition into a framework, then the elaboration of that framework into various dimensions, and then the development of test items under each dimension that conformed to the framework. These go through a range of processes – translation, field trials, etc. When, finally, the booklets with the test

⁵⁵ Nine of the countries took the paper-based assessment (PBA) and were not eligible to participate in the assessment of global competence, which was only designed as a computer-based assessment (CBA).

items are all assembled and printed off, they inscribe within them a particular notion of GC and a particular image of 15-year-olds that have been stabilized through great effort, after experts have flown many miles and conferred for many hours, and after objections have been raised and successfully quelled. But the trials of this concept of GC are not over yet. They are yet to meet the 15-year-olds who are the intended test takers for this test. In the next section, we explore what happens when the survey meets the students.

Things Fall Apart: Standardized Surveys Meet Non-Standard 15-year-olds

As mentioned previously, the test items developed for each PISA domain are based on assessment frameworks that are required to fulfil three main criteria: everyday relevance, domain specificity and appropriateness for the developmental level of 15-year-olds (Sälzer & Roczen 2018, p. 8). These criteria are used by experts to assess PISA frameworks, based on their assumptions about the lives of distributed, diverse 15-year-olds. The student questionnaires had been field-trialed, so it had been vetted by real 15-year-olds. This was not the case with the cognitive surveys. Our method of survey encounters performed a dual function – on the one hand, it provided a useful and effective way to elicit insights into the students' thinking which could not be captured in their multiple-choice responses. At the same time, it also provided an opportunity to the test whether the assumptions about their lives inscribed into the questionnaire were reasonable.

For the survey encounters, Author #1 prepared a curated sample of questions from the open access student questionnaires. The questions presented to the students were taken from each of the 15 constructs that made up the self-reported survey. The students, in groups of three, were clustered around one copy of the survey questions given to them. The students were instructed to discuss each question, collaboratively agree on a response, and fill it in. Their discussions were audio recorded. This was followed up with interviews where the students were asked to provide further explanation to their responses. Below are some examples of how the students read and responded to the survey.

Reducing Energy

One aspect of PISA's definition of a globally competent individual is the extent to which they are able 'to act for collective well-being and sustainable development' (OECD 2018, p. 7). Translating this aspect into the survey, students are asked to report on their engagement with global issues. 'Global issues' have been defined by PISA as issues which 'affect all individuals, regardless of their nation or social group' (OECD, 2018, p.12). Examples of these issues include, among other things, energy use at home, choosing to buy more ethical or environmentally sustainable products, and boycotting products for political, ethical or environmental reasons. These issues, according to PISA, would be relevant to all 15-year-olds, thus meeting the first criterion. The assumption is that a 'globally competent' 15-year-old is one who is engaged with all these issues and will respond with a 'Yes' to all the statements. However, the example below shows that these issues are too complex to be easily answered with a binary yes/no. There are many levels of nuance embedded within the students' responses.

In this question, a statement was presented, and the students had to report if the statement described them or not.

I reduce the energy I use at home (e.g. by turning the heating down or turning the air conditioning up or down or by turning off the lights when leaving a room) to protect the environment. (OECD, 2018, p. 55)

Faced with this question, Charles responded:

So... my mom's pretty strict about turning the heating down... keeping the energy use at like a certain point [because] she's really conscious about how much money we spend on our electricity bills and our water bills. So, it's kinds of not for the purpose of protecting the environment, more from an economic... standpoint. But then I guess it's something that's become a force of habit. So yes.

While the question is looking to obtain the students' attitude towards protecting the environment, Charles states his actions are motivated by his mother's interest in reducing their bills. In this example, rather than 'acting for collective well-being and sustainable development' (OECD, 2018, p. 7), as intended by PISA, Charles response of 'Yes' on the survey reflects his compliance to his parents. This response will show Charles as an individual who engages in 'tackling... environmental challenges' and supporting the SDGs (OECD, 2018, p. 5) while the survey encounter does not support that result. To this same question another student, Cody, stated:

I'll say that... it's a yes. I'm better at some things than others. So, for example, I take pretty long showers which wastes water I guess... but I'm also pretty good at like turning the lights off, turning the heater off and all that which saves electricity... but I try my best.

In this question, students are assumed to either be fully committed to reducing the energy at home or not at all. The answer options presented do not give them a chance to choose a middle ground. Cody's response highlights one of the issues inherent in most self-reported surveys: given only an option of yes/no more respondents who were in the 'middle ground' veered toward the more socially desirable answer (Nicolaas, Campanelli, Hope, Jäckle, & Lynn, 2015). This aspect has been acknowledged by the OECD as a 'common problem' in self-reported attitudinal surveys (OECD, 2018, p. 34).

Embedded in the above survey item is the assumption that all test takers have heating or air-conditioning – or indeed electricity. This reflects the analysis of Ledger et al. (2019) that the constructs in this assessment:

essentially describe the habitus of a global elite, making it hard to see how a child from a lower socioeconomic background and/or an attendee of a poorly funded local school could possibly score well on this scale. (p. 19)

For students in a low SES school in a low-income nation, where electricity is not guaranteed and the luxuries of heating and air conditioning are absent (or who live in climates where these are not required), this question might not meet the criterion of relevance to everyday situations. Moreover, just as Serder and Jakobsson (2015) had found in their investigations, here, too, the students' responses did not necessarily elicit the information the survey sought to elicit.

Choosing Ethical or Environmentally Friendly Products

Another statement to assess the students' engagement with global issues is: *I choose certain products for ethical or environmental reasons, even if they are a bit more expensive* (OECD, 2018, p. 55). As with the previous example, this item could not be easily answered with just a yes/no. According to Marilyn:

Depends... I don't... buy a lot of my own stuff so sometimes I don't necessarily have a choice but also like when I am buying my own stuff sometimes, I can't really afford the... more expensive thing.

Echoing Marilyn, Marigold states, 'I would, but I am poor'. Both Marilyn and Marigold then check No in response to this item.

The questions in this item carry a complex interweaving and overlapping of knowledge, attitudes and behaviors. While the students had said that they would buy ethical products (attitude), because, based on their studies, they knew that it was the 'right' thing to do (knowledge), they were unable 'to act for collective well-being and sustainable development' (OECD, 2018, p. 7) (behavior). Examples from this survey encounter support Serder's (2019) findings that what 'might seem a general-life situation to the test constructors' may not uniformly represent 15-year-olds across the globe (Serder, 2019, p. 81). Taken in the context of Australia, this question does not fulfil the criterion of the assessment framework, which requires that the question is 'appropriate for capturing the developmental stage of 15-year-olds' (Sälzer & Roczen 2018, p. 8). This point is clearly illustrated by Megan:

...because I am a high school student, I don't have that much responsibility in the world yet. I am still a minor. First of all, I can't really speak for myself legally or anything like that yet. So, I know I can't really do that much. But I would like to actively participate, helping people in other countries, but I know that there are many cons of doing that... it is not always sustainable or anything, but I would like to do something.

Thus this 'hybrid of cognitive components and personal attitudes of the assessed students' (Sälzer & Roczen 2018, p. 10) is not adequately captured by the various dimensions of PISA's GC survey. The vision of a standardised, generic and homogeneous group of 15-year-olds that appears to be imagined by the item developers splinters when confronted with the diverse, complex, thoughtful students of the real world.

Global Mindedness

PISA's conceptualization of global mindedness came from a number of sources including the Global-Mindedness Scale developed by Jane Hett in 1993 (OECD, 2018, p. 35). Building on existing conceptualizations of what a globally minded individual looks like, the survey includes six questions that seek to assess the students' *sense of world citizenship*, *responsibility for others in the world*, *sense of inter-connectedness*, and *global self-efficacy* (OECD, 2018, p. 36). Below are responses to one of these items: *I think of myself as a citizen of the world* (OECD, 2018, p. 58). Students were required to respond to these items by ticking one of the following: *strongly disagree*, *disagree*, *agree*, or *strongly agree*.

A 'citizen of the world', according to the OECD, is an individual who has 'commitments and obligations toward the planet and others, irrespective of their particular cultural or national background' (OECD, 2018, p. 17). While the OECD has a clear view of what they mean by this term, the survey encounters revealed that a number of students were unclear as to what this concept meant. For example, Marilyn, in an attempt to understand what a 'citizen of the world' meant, commented:

Maybe if you are a citizen of that country you belong to that country, so it means that you belong as part of the world? I don't know.

Marilyn checked the *Agree* box.

Stephanie revealed a different understanding of what it means to be a citizen of the world:

Being a human being in an accepted... being able to be yourself as a human being in the world and not having to... prove yourself. I don't know, I just feel like you should be accepted as a human being that's what like makes a citizen.

Based on this understanding of the phrase 'citizen of the world', Stephanie checked *Strongly agree*. Ticking an answer without clearly understanding the question has implications for the validity and reliability of the results and it would ultimately impact on the assessment's ability to inform policy changes.

One criterion for developing the test constructs is that the items appropriately capture the developmental stage – i.e., based on the response, students' level of GC can be graded. From the examples above it is not clear whether the items fulfil the criteria. As they did not understand the question, their developmental level cannot be readily inferred from their responses.

Sälzer and Roczen (2018) explain that the development of clear coding (i.e., marking or grading) guidelines are dependent on clear definitions and well-developed theory, as well the ability to link questions to difficulty levels or responses to proficiency levels. For literacy and numeracy, these guidelines have come to be well developed. For GC, they argue, such classifications and gradations are more difficult to establish. They add:

For many of the test units developed to assess global competence and proposed by the OECD, however, it was not even obvious to National Project Managers how the degree of correctness of responses was to be derived and how correct responses differed from less correct or wrong ones. Delimiting correct from incorrect responses is, accordingly, strongly dependent on the test-takers' presuppositions, so that a consensus seems to be inconceivable at this point. (Sälzer & Roczen, 2018, p. 12)

Emphasizing the importance of theory for the development of useful questions and clear coding guidelines, Sälzer and Roczen state that the development of theory should have preceded the development of the assessment surveys. Although their remarks pertain to the 2016 frameworks, our analysis shows that they apply equally to the 2018 iteration:

The assessment framework for global competence (OECD, 2016c), which was established over only a few months, could not compensate for a lack of thorough groundwork on the concept. The criterion of selecting central important theories and concepts for each domain in PISA (OECD, 2016b) is therefore only insufficiently met for global competence. This is also reflected in the fact that, at different stages of development in which the framework was made accessible to those in charge of running the PISA study in the different countries, the definition and elements of global competence varied fundamentally. As a result, the framework gave the impression of being rather arbitrarily designed and the inadequacy of a theoretical consolidation of the construct became apparent. (Sälzer & Roczen, 2018, p. 13)

Taken together, the survey encounters revealed that the test items were unable to consistently meet some of the criteria. The stability that was attained when the survey instruments were finally assembled, somewhat late, and without full field trial, proved to be fleeting – it began to unravel as real-life 15-year-olds (as opposed to those imagined by the test developers) encountered the survey. In some cases, the students arrived at an answer based on criteria that were different to what the OECD sought to assess. In others, the scenarios were not appropriate to the development

levels of the students. In yet others, students' responses were more interesting and complex than the questions accommodated.

Conclusions

In this paper, we took a sociology of quantification approach to determine whether the PISA survey was a robust enough instrument to usefully measure GC and to provide useful information to policy makers, curriculum designers and teachers, as it seeks to do. Rather than taking on the role of sole evaluators of the test items, we traced the process by which this survey was assembled, taking note of the different objections raised and noting how, eventually, the survey was able to convince the PGB enough to get the permission to proceed with the test, even though several PISA participants opted out of the survey.

One question of interest related to the assumptions about GC that were inscribed in the assessments. We found that there was a clear, Western or global north bias in the scenarios implied in the questionnaire. To some extent, since our survey encounters were also conducted in the 'global north', this did not become as evident as it might have, had we done our survey encounters in, for example, Sub-Saharan Africa.

Sälzer and Roczen (2018) have noted, based on the 2016 versions of the instruments, that the scenarios presented in the assessment promoted stereotypes, and that some items were abandoned after receiving objections. The PISA 2018 document provides glimpses of the stereotypical thinking of the GEG itself, when it attributes certain traits to certain cultures. Here is one example:

people from some cultural backgrounds tend to exaggerate their responses to typical questionnaire items based on a Likert-type scale (e.g. questions asking students whether they strongly disagree, disagree, agree or strongly agree with a statement), whereas others tend to take a middle ground (Harzing, 2006, as cited in OECD, 2018, p. 22)

Such linking of 'people from some cultural backgrounds' (rather than just 'some people') with particular behaviors is indeed the definition of racist thinking. It is most ironic that this statement appears in a document that seeks to assess 'global competence'. It could be argued that the OECD is basing this statement on 'research evidence' and is not making up this idea on its own – but the ability to discriminate between good and bad evidence, and between responsible and irresponsible communication, is one of the attributes that the GC instruments seek to assess.

The PISA GC team acknowledges some of the limitations inherent in assessing such phenomena as 'global competence' – but only in relation to the student questionnaire and not the cognitive assessment:

For some of the questions measuring attitudinal or socio-emotional traits (e.g. "openness"), however, defining right or wrong answers is more controversial because the development of these traits and their contribution towards global competence might be non-linear (beyond a certain threshold, more "openness" may not necessarily be better). (OECD, 2018, p. 22)

PISA's solution to this issue is to not use the student questionnaire results in preparing country rankings but use them only to 'illustrate general patterns and differences within countries ... as well as to analyze the relationship between those skills and attitudes and students' results on the cognitive test' (OECD, 2018, p. 22). Given that the construct underpinning both parts of the survey

is so underdeveloped, the conclusions that will be drawn and the ‘best practices’ derived will be scientifically unsound, but may nevertheless influence policies, curricula and pedagogies.

Our paper adds to the excellent political commentary and empirical analyses that already exists of this issue. We take a socio-technical approach in which the strength of the instrument is not assumed to be inherently about whether it closely models the phenomenon, since the phenomenon itself is produced by such instruments. Rather, we explore the technical work (eliminating certain questions, using only the cognitive survey in the rankings) and the political work (changing the GEG team, linking with the SDGs) that allowed the OECD to overcome the objections of enough nations to proceed with the assessments. The survey encounters provide an insight into the ways in which one set of 15-year-olds thought about the issues that are deemed to be part of GC. Even though our encounters were performed in the global north (and the instruments appeared to carry a global north bias), our 15-year-olds could not readily relate to some of the situations and questions. The questions failed also to capture the complexity of students’ lives and thinking, and their responses could not express their experiences or attitudes.

The PISA 2018 assessment of GC can be viewed in the context of the larger issue of the ambitions of global agencies to ‘fix’ education globally, based on global assessments and comparisons and global monitoring instruments. The limitations placed by the need to create internationally standardized instruments on the authenticity of assessment – even in relation to well-theorized entities such as numeracy – have been well documented (see, for example, Hutchison & Schagen, 2007). For constructs such as GC, this limitation is greatly increased – not only because the concept is undertheorized, but also because what counts as ‘GC’ is so context-specific that a ‘universal standard’ is neither realistic nor desirable. Indeed, the idea of measuring ‘global competency’ through a globally standardized instrument is paradoxical, and the attempt to do so is ironic.

References

- Addey, C. (2019). Researching inside the international testing machine: PISA parties, midnight emails & red shoes. In B. Maddox (Ed.), *International large-scale assessments in education. Insider research perspectives* (pp. 13-29). Bloomsbury Publishing.
- Auld, E., & Morris, P. (2019). Science by streetlight and the OECD’s measure of global competence: A new yardstick for internationalisation? *Policy Futures in Education*.
<https://doi.org/10.1177/1478210318819246>
- Chandir, H. (2020). Student responses on the survey of global competence in PISA 2018, *Discourse: Studies in the Cultural Politics of Education*. <https://doi.org/10.1080/01596306.2020.1844153>
- Council of Europe. (2016). *Competences for democratic culture. Living together as equals in culturally diverse democratic societies*. Council of Europe Publishing.
- Deardorff, D. K. (2006). Identification and Assessment of Intercultural Competence as a Student Outcome of Internationalization. *International Education*, 10(3), 241-266.
<https://doi.org/10.1177/1028315306287002>.
- Derksen, L. (2000). Towards a sociology of measurement: The meaning of measurement error in the case of DNA profiling. *Social studies of science*, 30(6), 803-845.
<https://doi.org/10.1177/030631200030006001>
- Desrosières, A. (1998). *The politics of large numbers*. (C. Naish, Trans.). Harvard University Press.
- Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, 113(1), 1-40. <https://doi.org/10.1086/517897>

- Gorur, R. (2011). ANT on the PISA trail: Following the statistical pursuit of certainty. *Educational Philosophy and Theory*, 43(sup1), 76-93. <https://doi.org/10.1111/j.1469-5812.2009.00612.x>
- Gorur, R. (2014). Towards a sociology of measurement in education policy. *European Educational Research Journal*, 13(1), 58-72. <https://doi.org/10.2304/eerj.2014.13.1.58>
- Gorur, R. (2015). Producing calculable worlds: Education at a glance. *Discourse: Studies in the Cultural Politics of Education*, 36(4), 578-595. <https://doi.org/10.1080/01596306.2015.974942>
- Gorur, R. (2017). Towards productive critique of large-scale comparisons in education. *Critical Studies in Education*, 58(3), 341-355. <https://doi.org/10.1080/17508487.2017.1327876>
- Grotlüschen, A. (2018). Global competence—Does the new OECD competence domain ignore the global South? *Studies in the Education of Adults*, 50(2), 185-202. <https://doi.org/10.1080/02660830.2018.1523100>
- Haas, P. M. (1992). Introduction: Epistemic communities and international policy coordination. *International organization*, 46(1), 1-35. <https://doi.org/10.1017/S0020818300001442>
- Hacking, I. (1990). *The taming of chance* (Vol. 17): Cambridge University Press.
- Hardy, I. (2015). 'I'm just a numbers person': The complexity, nature and effects of the quantification of education. *International Studies in Sociology of Education*, 25(1), 20-37. <https://doi.org/10.1080/09620214.2014.972971>
- Hutchison, D., & Schagen, I. (2007). Comparisons between PISA and TIMSS: Are we the man with two watches. In T. Loveless (Ed.), *Lessons learned: What international assessments tell us about math achievement* (pp. 227-261). Brookings Institution Press.
- International Baccalaureate Organisation. (2015). *What is an IB education?* <http://www.ibo.org/globalassets/digital-toolkit/brochures/what-is-an-ib-education-en.pdf>
- Jasanoff, S. (2004). Ordering knowledge, ordering society. In S. Jasanoff (Ed.), *States of knowledge: The co-production of science and social order* (pp. 13-45). Routledge.
- Knorr-Cetina, K. (1981). *The manufacture of knowledge*. Pergamon.
- Knorr-Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*: Harvard University Press.
- Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*: Harvard University Press.
- Le Hebel, F., Montpied, P., & Tiberghien, A. (2014). Which effective competencies do students use in PISA assessment of scientific literacy? In C. Bruguiere, A. Tiberghien & P. Clement (Eds.), *Topics and trends in current science education* (pp. 273-289): Springer.
- Ledger, S., Thier, M., Bailey, L., & Pitts, C. (2019). OECD's approach to measuring global competency: Powerful voices shaping education. *Teachers College Record*, 121(8), 1-40.
- Mol, A. (1999). Ontological politics. A word and some questions. *The Sociological Review*, 47(1_suppl), 74-89.
- Nicolaas, G., Campanelli, P., Hope, S., Jäckle, A., & Lynn, P. (2015). *Revisiting "yes/no" versus "check all that apply"*: Results from a mixed modes experiment. [Paper presentation]. Survey Research Methods. <https://doi.org/10.18148/srm/2015.v9i3.6151>
- OECD. (2014). *PISA 2018 Framework plans. 38th meeting of the PISA Governing Board*. [https://one.oecd.org/document/EDU/PISA/GB\(2014\)16/en/pdf](https://one.oecd.org/document/EDU/PISA/GB(2014)16/en/pdf)
- OECD. (2016). *Global competency for an inclusive world*. <https://www.oecd.org/education/Global-competency-for-an-inclusive-world.pdf>
- OECD. (2018). *Preparing our youth for an inclusive and sustainable world: The OECD PISA global competence framework*. www.oecd.org/pisa/Handbook-PISA-2018-Global-Competence.pdf
- Poovey, M. (1998). *A history of the modern fact: Problems of knowledge in the sciences of wealth and society*. University of Chicago Press.

- Porter, T. M. (1996). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton University Press.
- Reimers, F. M. (2013). *Assessing global education: An opportunity for the OECD*. <http://search.oecd.org/pisa/pisaproducts/Global-Competency.pdf>
- Sälzer, C., & Roczen, N. (2018). Assessing global competence in PISA 2018: Challenges and approaches to capturing a complex construct. *International Journal of Development Education and Global Learning*, 10(1), 5-20. <https://doi.org/10.18546/IJDEGL.10.1.02>
- Sandoval-Hernández, A., Isac, M. M., & Miranda, D. (2019). *Measurement strategy for SDG Global Indicator 4.7.1 and Thematic Indicators 4.7.4 and 4.7.5 using international largescale assessments in education proposal*. <http://gaml.uis.unesco.org/wp-content/uploads/sites/2/2019/08/GAML6-REF-9-measurement-strategy-for-4.7.1-4.7.4-4.7.5.pdf>
- Sauvageot, C. (2008). A tool for international comparisons: The International Standard Classification of Education (ISCED). *Education & Formations: International comparisons*, 211-221. https://cache.media.enseignementsup-recherche.gouv.fr/file/Publications/75/6/EDUC&FORM_EN_41756.pdf#page=210
- Schleicher, A. (2016). Pisa tests to include 'global skills' and cultural awareness. *BBC News*. Retrieved from <https://www.bbc.com/news/business-36343602>
- Sellar, S., & Lingard, B. (2014). The OECD and the expansion of PISA: New global modes of governance in education. *British Educational Research Journal*, 40(6), 917-936. <https://doi.org/10.1002/berj.3120>
- Serder, M. (2019). Detecting student performance in large-scale assessments. In B. Maddox (Ed.), *International large-scale assessments in education: Insider research perspectives* (p. 69). Bloomsbury Publishing.
- Serder, M., & Jakobsson, A. (2015). "Why bother so incredibly much?": Student perspectives on PISA science assignments. *Cultural Studies of Science Education*, 10(3), 833-853. <https://doi.org/10.1007/s11422-013-9550-3>
- Skirbekk, V., Potancokova, M., & Stonawski, M. (2013). *Measurement of global citizenship education*. Study commissioned by UNESCO for the technical consultation on GCE in Seoul, Republic of Korea. https://www.gcedclearinghouse.org/sites/default/files/resources/Measurement%20of%20global%20citizenship%20education_0.pdf
- UNESCO. (2013). *Outcome document of the technical consultation on global citizenship education*. <https://unesdoc.unesco.org/ark:/48223/pf0000224115>
- UNESCO. (2014). *Global citizenship education: Preparing learners for the challenges of the 21st century* <http://unesdoc.unesco.org/images/0022/002277/227729E.pdf>
- UNESCO. (2016). *Unpacking Sustainable Development Goal 4: Education 2030*. <https://unesdoc.unesco.org/ark:/48223/pf0000246300>
- Voogt, J., & Roblin, N. P. (2012). A comparative analysis of international frameworks for 21st century competences: Implications for national curriculum policies. *Journal of Curriculum Studies*, 44(3), 299-321. <https://doi.org/10.1080/00220272.2012.668938>
- Woolgar, S. (1991). Beyond the citation debate: Towards a sociology of measurement technologies and their use in science policy. *Science and Public Policy*, 18(5), 319-326. <https://doi.org/10.1093/spp/18.5.319>

About the Authors

Harsha Chandir

Deakin University

h.chandir@deakin.edu.au

<https://orcid.org/0000-0002-3931-3296>

Harsha Chandir, PhD., is a research fellow at Deakin University. Her areas of research include international large-scale assessments, education policy and curriculum design.

Radhika Gorur

Deakin University

radhika.gorur@deakin.edu.au

<http://orcid.org/0000-0002-4528-0793>

Radhika Gorur, PhD., is an Associate Professor at Deakin University, and a Director of the Laboratory of International Assessment Studies. Currently, with a grant from the Australian Research Council, she is researching the role of global policy networks and the new accountability practices in the Indo-Pacific, with an empirical focus on Cambodia and India to develop principles for sustainable, participatory accountability practices.

About the Editors

Oren Pizmony-Levy

Teachers College, Columbia University

op2183@tc.columbia.edu

Oren Pizmony-Levy is an associate professor in the Department of International and Transcultural Studies at Teachers College, Columbia University. He holds a PhD in sociology and comparative and international education from Indiana University-Bloomington. His scholarship focuses on the intersection between education and social movements. Through diverse set of research methods, he studies the roots and impact of global educational movements – including international large-scale assessments, environmental and sustainability education, and LGBT education. He is the founding director of the Center for Sustainable Futures and a co-leader of the New York City Partnership for Sustainability Education, a research-practice partnership between the Department of Education and the Center.

Dafna Gan

Kibbutzim College of Education, Technology, and the Arts

dafna.gan@gmail.com

Dr. Gan, Ed.D, an environmental and sustainability education researcher and lecturer at the Kibbutzim College of Education, Technology, and the Arts, Israel. She is the program director of the Environmental Education Master program at the science department. Her research focuses on environmental and sustainability education, multicultural education, transformative learning, leadership in higher education and non-governmental sustainability organizations. Her academic practices are devoted to sustainability implementation in both the college sector and in the educational system in Israel. She is the founder and director of the Center for Sustainability Education - Integrating Environmental Educational Research, and Practice.

SPECIAL ISSUE
**Learning Assessments for Sustainability? Exploring the Interaction
between Two Global Movements**

education policy analysis archives

Volume 29 Number 122

September 27, 2021

ISSN 1068-2341



Readers are free to copy, display, distribute, and adapt this article, as long as the work is attributed to the author(s) and **Education Policy Analysis Archives**, the changes are identified, and the same license applies to the derivative work. More details of this Creative Commons license are available at <https://creativecommons.org/licenses/by-sa/4.0/>. **EPAA** is published by the Mary Lou Fulton Institute and Graduate School of Education at Arizona State University. Articles are indexed in CIRC (Clasificación Integrada de Revistas Científicas, Spain), DIALNET (Spain), [Directory of Open Access Journals](#), EBSCO Education Research Complete, ERIC, Education Full Text (H.W. Wilson), QUALIS A1 (Brazil), SCImago Journal Rank, SCOPUS, SOCOLAR (China).

About the EPAA/AAPE Editorial Team:

<https://epaa.asu.edu/ojs/index.php/epaa/about/editorialTeam>

Please send errata notes to Audrey Amrein-Beardsley at audrey.beardsley@asu.edu

Join EPAA's Facebook community at <https://www.facebook.com/EPAAAPE> and Twitter feed @epaa_aape.