Assessment of the generic problem-solving construct across different contexts

A K L Nguyen¹, C T K Nguyen¹, and R J Adams¹,²

¹Assessment Research Centre, The University of Melbourne, Vic 3010, Australia
²Australian Council for Educational Research, VIC 3124, Australia

E-mail: lananh289@gmail.com

Abstract: Problem-solving, one of the generic skills in the so-called ‘21st century’ skillset, has been found to be critical to the personal and professional success of learners. This study evaluates the existence of conceptual meaning of students’ generic problem-solving (GPS) competency in three distinct contexts, namely mathematics, reading, and creative problem-solving using the PISA Computer-Based Assessments (CBA) 2012 data. The analyses were framed in the context of a content validity exploration from Messick’s validation framework to reveal the GPS conceptual meaning underlying PISA CBA 2012 measurements. Specific tasks carried out include: (a) conducting skill auditing of PISA CBA 2012 items to understand their measured skill content and relevant to and representative of the generic problem-solving construct’s level; (b) using the outcome of the skill auditing stage to interpret the GPS competency levels measured by the domain-specific subject tests before comparing them with corresponding levels together.

1. Background and objective of the study
Over the last 40 years, economic and technological changes have drastically reshaped education systems throughout the world. Problem-solving (PS) is arguably one of the most pertinent due to its inherent ability to deal with complexity and its mutual relationship with other skills. Numerous studies have been conducted worldwide to assess students’ PS in a specific range of contexts. However, none of these studies has specifically identified PS competencies that are likely to transfer across more than three different PS contexts. This leads to the question of whether there is a generic problem-solving (GPS) construct that can apply in a wide range of fields. The answer to this question has not been adequately justified. This study investigated the extent to which a single GPS construct can be invariant across mathematics, reading and creative PS. The objectives of the research were (a) to identify the measured skills in PISA 2012 computer-based assessments (CBA) relevant to and representative of the GPS construct’s level and (b) to compare the interpretations of the measured GPS competency levels in the domain-specific areas together.

2. Theoretical framework
In educational assessment, the assessment for a GPS construct’s existence across fields has received a lot of interest due to an increased emphasis on preparing students for 21st century skills. A shared interest for researchers, policymakers, educators and the public is whether a GPS construct exists across different domain-specific areas and can thus be taught in the classroom [1]. Instances of students performing similarly in solving problems in different contexts, for example, in analytic PS and
interactive PS [2], and the existence of GPS skills which are independent of any specific context, make imperative the need for assessing the existence of a GPS construct. To explore the existence of a GPS construct, most studies have evolved by comparing different PS constructs to reveal the similarities and differences between those constructs [1, 2]. However, it seems impossible to find any valid research in the literature focusing only on a comparison of three or more measured constructs in PS tests.

This paper intends to provide a validity judgement approach by adapting the content validity process from Messick [3]’s validation framework to reveal the GPS conceptual meaning measured by PISA CBA 2012 instruments. Messick [3] acknowledged the importance of expert judgement to evaluate content relevance, by which researchers look for evidence of content relevance and representativeness of the tests. Specifically, to justify the claim that GPS skills underlie three domain-specific PS constructs, the evaluation of this claim would begin with assessing the evidence that the PISA 2012 item test content is relevant to description of GPS construct by using skill auditing. This stage leans on the skill-auditing process, which uses expert review to support judgements of the relevance and representativeness of the content upon which items are based. A skill-auditing process has been applied to the content validity in different studies. For example, Tran, Griffin and Nguyen [4] undertook a skill-auditing process as the evidence for content validity on the interpretation and use of the 2008 University Entrance Examination English test scores.

The research questions to be addressed in this paper are as follows:

- To what extent are the skills that underlie the PISA CBA 2012 released items relevant to and representative of the levels in the GPS construct?
- To what extent do the interpretations of item ordering in the construct of GPS competence measured by the PISA CBA 2012 released items match with each other?

3. The secondary PISA CBA 2012 data and the participants

This study used both previously collected data and direct correspondence from human participants. The secondary data, including released items and analysis reports from PISA CBA 2012, were downloaded from the OECD website (www.oecd.org/pisa/pisaprod/database-cbapisa2012.htm). Data from the participants of 32 countries/economies, with 117,933 students, were used in this study. These released computer-based items can be seen at www.oecd.org/pisa or http://erasq.acer.edu.au (username: public; password: access). No new reading material was released after the 2012 survey administration. There were 30 released items including both multiple-choice and partial-credit formats. There are two reasons why this study focused only on computer-based PISA 2012 data. First, for an equal comparison between constructs, the method of delivering the tests needs to be the same in three assessments. The interactive PS assessment was administered in a computer environment. As a result, the computer-based mathematics test and the digital reading test in PISA 2012 were chosen for a similar administered approach. There were also a science test and a financial test in the PISA 2012 cycles; however, they were only administered in paper-and-pencil form. Thus, neither was included in this research. Second, a requirement of this study was that one student had to answer three assessments at a similar time. Thus, only PISA 2012 CBA data could satisfy this condition.

In collecting subjects' experts at the skill-auditing stage, the targeted participant group comprised six adults 22 years or older who were competent to give consent. They were senior testing developers and specialists in the field of testing, assessment and evaluation in Australia. They were also selected based on their familiarity with the PISA tests, mathematics assessment, reading assessment and PS assessment. Individual opinions from the experts regarding the skills needed to answer the questions in the PISA CBA 2012 tests were collected via an item analysis form. To record the discussion, as well as using participants’ completed forms, the researcher employed both audio recording and note-taking. These data and materials were collected at the University of Melbourne site.
4. Research design

4.1. Using individual analysis and group discussion to explore the skill measured by the PISA 2012 computer-based released items relevant to and representative of the levels in the GPS competence

This analysis stage adopted the skill-auditing procedure suggested by Tran et al. [4] as this would provide high methodological clarity and the fundamental developmental PS (DPS) construct [5-7] as the reference framework. The description of the six levels of DPS is given in [5]. Like many taxonomies, the levels in this framework display a hierarchy in which learners have more probability of displaying the performance at a higher level as their PS competence increases. The purpose of this analysis was to describe the measured continuums in term of GPS skills and levels based on the item’s intended measured skills so as to compare the derived continuums’ descriptions in the next analysis. Skill auditing was evaluated in three different content areas, mathematics, reading, creative PS, varying in knowledge content. Thereby, the skill auditing of each item and level description of the GPS in each domain were analysed in depth.

4.2. Comparing the interpretations of item ordering in terms of GPS skills measured by the PISA 2012 computer-based tests

The second research question aimed to compare the interpretations of the GPS constructs measured by the PISA CBA 2012 released items to evaluate the extent to which the three derived constructs were similar. In this study, the similarities represented the existence of a GPS construct underlying three domain-specific PS constructs. To arrive at such an interpretation for the measured GPS constructs in three PISA CBA 2012 tests, the expert judgement-based approach was used. As Mach, Mastrandrea, Freeman and Field [8] pointed out, applying rigorous judgement is an important dimension of assessment. The results of the previous skill auditing helped to inform whether there was a common substantive interpretation of the underpinning construct. The released PISA CBA 2012 items were placed in descending order of levels, according to the experts’ judgement. The set of items in each level was then examined to interpret different measured GPS levels using the fundamental DPS proficiency and referring to the European Qualifications framework for generic competencies. Then, to validate these substantive interpretations, the researcher emailed these level descriptions to the expert participants and asked them to judge if these levels were meaningful, informative and useful. Finally, the interpretations of the GPS levels measured in PISA CBA 2012 released item tests were compared with the descriptions of corresponding levels. The purpose was to inform the similarities and differences between the three measured GPS levels in three domain-specific areas.

5. Results and discussion

5.1. The skill measured by PISA CBA 2012 released items relevant to and representative of the levels in the GPS construct

At the skill-auditing stage, expert participants identified the GPS skills measured by each PISA 2012 CBA test item (domain-specific areas) and located the item in the GPS levels. Differences could be seen in their individual analyses, but these were few and were mainly due to the varying extent of detail in their descriptions rather than a fundamental dissimilarity in their perceptions of what GPS skills were needed to succeed. With many years of testing experience, all the participants responded elaborately to the researcher’s questions of clarification and came to relatively uniform conclusions. Based on their individual analyses, the expert participants joined the group discussion to express their views. The judgement they provided at this stage was generally consistent with the points they made in the completed individual forms. Figure 1 presents the relations between PISA CPS 2012 levels (dash line) versus the skill-auditing GPS level (solid line) of released items across creative PS, computer-based mathematics and digital reading domains. Two main findings can be drawn from the skill-auditing results.
Figure 1. The relative between PISA CPS 2012 level versus skill auditing GPS level of released items across creative problem-solving area.

Figure 2. The relative between PISA CPS 2012 level versus skill auditing GPS level of released items across computer-based mathematic area.

Figure 3. The relative between PISA CPS 2012 level versus skill auditing GPS level of released items across digital reading area.
In the first finding, all the experts asserted that while it was difficult to classify the PISA CBA 2012 released items into GPS levels, the test content was highly relevant to the GPS skills descriptions in the six levels of the fundamental GPS construct. These comments were reasonable, and it was fair to say that the PISA CBA 2019 items were developed to measure literacy competencies such as mathematics, reading, or science constructs rather than the GPS construct. This can be observed in figure 1, figure 2 and figure 3, in which most of the released items have two different levels according to the PISA CBA 2012 scale and skill-auditing results. An example of such items is as follows. For item CR017Q01, the participants found that students would need to locate explicitly provided information using a conventional symbol system and a text label. Thus, strictly speaking, the item should be classified under both reading competence and GPS competence. From their experience, though, the participants judged that candidates would be more challenged in reading than in the GPS. Hence, it was agreed that item CR017Q01, which was classified at Level 3 according to the PISA CBA 2012 scale, should be classified at Level 2 of GPS.

For items CP007Q02, CP002Q08, CP002Q06, CP025Q01, CP025Q02, CP038Q02 and CP038Q01 in released PISA CPS 2012 tasks, CR013Q04 and CR017Q07 in released PISA DRA 2012 tasks and CM015Q03, CM020Q02, CM020Q03, CM020Q04 and CM038Q06 in released PISA CBMA 2012 tasks, though expert participants’ detailed descriptions of the underlying knowledge and skills were similar, their individual level classifications differed significantly. Each item was classified under at least three or four levels. During the group meeting, they agreed that the classified level was the highest one, and success in these items required students to use most of the GPS skills at that level. Thus, these items were finally categorised under only one level (see figure 1, figure 2, and figure 3).

For items CR017Q07 full credit and CR017Q07 partial credit, especially, the expert participants explained that these items first tested students’ ability to explore the individual elements within the task in a random manner. Because these skills were classified under the GPS competence, two experts classified them under Level 2, two classified them under Level 4, one classified them under Level 3 and one classified them under Level 5. However, when they came to analyse item CR017Q07, they found that these items strongly tested students’ ability to recognise the message as potentially misleading or risky, and to evaluate the content of a text, which was classified under reading ability. After a long discussion, the experts agreed that items tested purely reading ability rather than GPS competence. Thus, the items CR017Q07 full credit and CR017Q07 partial credit were omitted from this study. This is why, in table 1, there is no solid line for these items when they belong to Level 4 and Level 3, according to the PISA CBA 2012 scale.

In the second finding, while the views on GPS skills relevance were highly complementary, those on level representativeness were not as positive. The results of skill auditing showed that the released PISA CBA 2012 items measured students in only three levels of GPS competence. On the one hand, the released items covered moderate GPS levels such as Level 3. It tested this level extensively, given that, of 30 items, 16 required students to explore and then recognise patterns. On the other hand, all the expert participants found that the measures for Level 6 and Level 1 were not represented by the released PISA CBA 2012 items. The measures of remainder of the levels (5, 4, and 2) were only limitedly in all of three domain-specific areas by several released items. Thus, the PISA CBA 2012 released test items were deemed able to measure a limited range of GPS competencies.

On a more positive note, the results of the inquiry showed that the released test items and the marking key were appropriate for rendering evidence of students’ GPS ability. All the experts acknowledged that the PISA CBA 2012 released items were in close accordance with the GPS skills. They asserted that, though these PISA CBA 2012 tests more generally measured the literacy abilities, students needed to have skills and abilities related to GPS competence to answer the items correctly. During the discussion, all the experts agreed with the skills needed to answer items correctly. The expert participants further shared that they and other colleagues had developed these PS tasks in several areas such as mathematics, reading or social science. Therefore, according to the experts, the test tasks were relevant to the fundamental GPS construct.
5.2. Comparing the interpretations of item ordering in terms of GPS skills measured by the PISA 2012 computer-based tests

To interpret the measured GPS levels in the PISA CBA 2012 tests and compare these interpretations with the descriptions of corresponding levels, the fundamental GPS construct was used as the reference framework. This study also referred to the illustrative descriptors from eight levels of the European Qualifications framework for generic competencies, elaborating descriptive progressions from UNESCO [9] to synthesise elaborated descriptors for the dimensions of the GPS skills to increase the amount of information for interpretive purposes.

Table 1 presents the interpretation of GPS levels in three domain-specific areas. As can be seen from table 1, the GPS interpretations in three domain-specific areas were compared with corresponding levels. The GPS Level 1 had no data in all three domains, and Level 6 was not present in mathematics and reading because there were no data in these domains. Beyond those mismatches, there were many similarities across the rest of the corresponding GPS levels in the three domains.

For example, the descriptions of GPS Level 2 (explore patterns) in the mathematics domain were equal to the descriptions of this level in the reading domain. Students in this low level in both scales had mastered the basic knowledge of domains and were starting to explore, search, locate, retrieve and interpret information. There were no data at this level in the CPS domain.

The similarities between GPS levels across three domains were also revealed in the comparison between Level 3, Level 4 and the highest Level 5 (testing rules and forming hypotheses). The highest level in the three areas involved formulating rules by identifying the relevant factors in the problem and their interrelationships, formulating rules to solve the problem according to the given constraints, and synthesising information and forming an opinion with evidence to support reasons. In other words, those students who could represent the rules were considered to have the highest GPS competency in all specific domains. Due to these commonalities, it can be concluded that the hierarchies of released items in the three domain-specific areas greatly resemble each other, which is another indicator that GPS competence exists across three domain-specific areas.

6. Conclusion and recommendations

Given the results presented and discussed so far, the findings for the study are clear. The content of the released PISA CBA 2012 items was relevant to the GPS construct’s levels. Nevertheless, since the released PISA CBA 2012 items targeted only a limited range of GPS skills, the released PISA CBA 2012 items had moderate representativeness of the measured GPS construct. On a brighter note, the results of the inquiry show that the released PISA CBA 2012 test items and the marking key are appropriate for rendering evidence of students’ GPS ability.

Given the experts’ rich experience of working with testing and assessment, their professional judgement presented so far has been an illuminating source of evidence. Because the four-level scales obtained were interpretable and meaningful, it is reasonable to conclude that there are some similarities in the score interpretations of the PISA CBA 2012 in the light of the GPS competency across three domain-specific areas. These were important pieces of evidence for the content aspects of validity. However, to achieve higher validity standards, more secured items should be included in future analyses as replacements for items within higher or lower GPS levels.

Special attention should be given to devising items that discriminate students in the ability range such as Level 6, Level 5, and Level 1 so that the competency of students in the GPS ability levels can be interpreted with greater precision. Izard [10] pointed out that items should also be designed in a way that they can be spread over different difficulty ranges so that evidence can be obtained for students of all ability levels.
Table 1. The GPS levels’ interpretations in three domain-specific areas measured by PISA CBA 2012 released items.

| GPS’s level based on experts’ judgments | Derived GPS levels measured by PISA CPS 2012 released items | Derived GPS levels measured by PISA CBMA 2012 released items | Derived GPS levels measured by PISA DRA 2012 released items |
|----------------------------------------|-------------------------------------------------------------|-------------------------------------------------------------|-------------------------------------------------------------|
| Level 5: Testing rules and forming hypotheses | Students at this level can organize and critically evaluate information, to find the (relevant) relationships between problem elements. This enables them to formulate complex hypotheses to predict the behaviour of the system. In controlling the execution of a solution strategy, they can plan ahead and monitor progress systematically to avoid errors and ensure actions contribute to the solution. | Students at this level can understand a mathematical concept and apply it in a new situation. They employ previous knowledge to form rules to solve the problem according to the given constraints. | Students at this level can access and retrieve information, combine, interpret, compare and evaluate a series of descriptions with a set of requirements. They synthesize information from several sources and form a conclusion supported by evidence and reasons. |
| Level 4: Forming rules | Students at this level can explore an unfamiliar system and identify patterns in different aspects of its behaviour. They can form and confirm a simple hypothesis as to the rules governing a system. | Students at this level can recognize and understand the relationship between sets of variables, to identify and reflect on a proposed incorrect solution. They execute a strategy to solve the task and give the correct solution. | Students at this level can use multiple steps to locate, search and retrieve required information. They evaluate, compare, integrate and interpret information and use the information to make a decision. They form rules, reflect on them and justify the suitability for a given situation with evidence. |
| Level 3: Recognizing patterns | Students at this level can explore a problem situation or system, recognise patterns in its components and infer simple relationships between these components. They can compare different plans or strategies and execute the best one. | Students at this level can recognize and understand relationships/connections/patterns in data and calculate correctly within them. They can manipulate a representation and apply it successfully. They can critically evaluate a statement about a representation, identifying flaws in it. | Students at this level can explore, extract, and retrieve information to identify a common criterion or find some connection between elements in the task. |
| Level 2: Explore patterns | N/A | Students at this level can understand a new, simple mathematical concept and apply it, but only partially successfully. | Students at this level can use background knowledge to explore elements of the task to find information that is explicitly stated. They search for patterns and connections between elements and may find them by trial and error. |
The results of this research have also raised important points regarding the test design of these measured GPS tests. The PISA CBA 2012 tests seem to have the features of a proficiency test. While the content of the PISA CBA 2012 released tests covered areas of domain-specific literacy, they also tested some areas of GPS skills such as Level 3 to Level 5. Though the GPS skills tested were relevant to the fundamental GPS construct, the reality is that not all the levels had been tested by the released items. On the other hand, it must be acknowledged that the designers of PISA CBA 2012 tests had done a good job in assembling items to measure PS in different subject areas. Although it seems no explicit assumption of a latent trait with specific levels of GPS ability had been nominated by OECD, most of the literacy domains tend to measure this GPS ability.

References

[1] Baird L L 1983 Review of problem solving skills. *ETS Research Report Series* 1

[2] Fischer A, Greiff S, Wüstenberg S, Fleischer J, Buchwald F and Funke J 2015 Assessing analytic and interactive aspects of problem solving competency *Learn. Individ. Differ.* 9 pp 172–9

[3] Messick S 1989 Validity, ed R L Linn *Educational Measurement* (New York: American Council on Education) pp 13–103

[4] Tran H P, Griffin P and Nguyen C 2010 Validating the university entrance English test to the Vietnam National University: A conceptual framework and methodology *Proced. Soc. Behv.* 2 pp 1295–304

[5] Rahman S, Griffin P and Zang Z 2016 Development and validation of the construct of problem solving *Prof. Australian Association for Research in Education Conf.* 2016 (Melbourne, Australia: AARE)

[6] Callingham R and Griffin P 2000 Towards a framework for numeracy assessment. *3rd Annual Conf. of the Mathematics Education Research Group of Australasia Incorporated* pp 134-141

[7] Callingham R and Griffin P 2001 Beyond the basics: Improving indigenous students' numeracy. *24th Annual Conf. of the Mathematics Education Research Group of Australasia Incorporated* pp 128-136

[8] Mach K J, Mastrandrea M D, Freeman P T and Field C B 2017 Unleashing expert judgment in assessment *Sci. Tech. (Global Environmental Change)* 44 pp 1–14

[9] UNESCO 2015 *Level-setting and Recognition of Learning Outcomes*, ed J Keevy and B Chakroun (Paris, France: The United Nations Educational, Scientific and Cultural Organization)

[10] Izard J 2005 Overview of test construction *Quantitative Research Methods in Educational Planning* ed K N Ross (Paris: UNESCO International Institute for Educational Planning) pp 1–51