International Journal of Science Education

ISSN: 0950-0693 (Print) 1464-5289 (Online) Journal homepage: http://www.tandfonline.com/loi/tsed20

Evaluation of diagnostic tools that tertiary teachers can apply to profile their students' conceptions

Madeleine Schultz, Gwendolyn A. Lawrie, Chantal H. Bailey, Simon B. Bedford, Tim R. Dargaville, Glennys O'Brien, Roy Tasker, Christopher D. Thompson, Mark Williams & Anthony H. Wright

To cite this article: Madeleine Schultz, Gwendolyn A. Lawrie, Chantal H. Bailey, Simon B. Bedford, Tim R. Dargaville, Glennys O'Brien, Roy Tasker, Christopher D. Thompson, Mark Williams & Anthony H. Wright (2017): Evaluation of diagnostic tools that tertiary teachers can apply to profile their students' conceptions, International Journal of Science Education, DOI: 10.1080/09500693.2017.1296980

To link to this article: http://dx.doi.org/10.1080/09500693.2017.1296980

INTERNATIONAL JOURNAL OF SCIENCE EDUCATION, 2017
http://dx.doi.org/10.1080/09500693.2017.1296980

Evaluation of diagnostic tools that tertiary teachers can apply to profile their students' conceptions

Madeleine Schultz[a], Gwendolyn A. Lawrie[b], Chantal H. Bailey[b], Simon B. Bedford[c], Tim R. Dargaville[a], Glennys O'Brien[c], Roy Tasker[d], Christopher D. Thompson[e], Mark Williams[f] and Anthony H. Wright[g]

[a]School of Chemistry, Physics and Mechanical Engineering, Queensland University of Technology, Brisbane, Australia; [b]School of Chemistry & Molecular Biosciences, The University of Queensland, St. Lucia, Australia; [c]School of Chemistry, University of Wollongong, Wollongong, Australia; [d]Department of Chemistry, Purdue University, West Lafayette, IN, USA; [e]School of Chemistry, Monash University, Clayton, Australia; [f]School of Science & Health, Western Sydney University, Penrith, Australia; [g]School of Education, The University of Queensland, St. Lucia, Australia

ABSTRACT

A multi-institution collaborative team of Australian chemistry education researchers, teaching a total of over 3000 first year chemistry students annually, has explored a tool for diagnosing students' prior conceptions as they enter tertiary chemistry courses. Five core topics were selected and clusters of diagnostic items were assembled linking related concepts in each topic together. An ordered multiple choice assessment strategy was adopted to enable provision of formative feedback to students through combination of the specific distractors that they chose. Concept items were either sourced from existing research instruments or developed by the project team. The outcome is a diagnostic tool consisting of five topic clusters of five concept items that has been delivered in large introductory chemistry classes at five Australian institutions. Statistical analysis of data has enabled exploration of the composition and validity of the instrument including a comparison between delivery of the complete 25 item instrument with subsets of five items, clustered by topic. This analysis revealed that most items retained their validity when delivered in small clusters. Tensions between the assembly, validation and delivery of diagnostic instruments for the purposes of acquiring robust psychometric research data versus their pragmatic use are considered in this study.

ARTICLE HISTORY
Received 8 July 2016
Accepted 15 February 2017

KEYWORDS
Concept diagnostic instrument; general chemistry; psychometric analysis; ordered multiple choice

Introduction

Effective teachers find out what students know by asking them questions, but what are the right questions to ask? Based on the answers, how should a teacher respond in their practices to support learning? These factors inform the task process and outcome of classroom learning (Biggs, 1993). This process represents the teachers' application of their topic-

CONTACT Madeleine Schultz madeleine.schultz@qut.edu.au Science and Engineering Faculty, CPME, Queensland University of Technology, Brisbane 4000, Australia

Supplemental data for this article can be accessed at: http://dx.doi.org/10.1080/09500693.2017.1296980.

specific professional knowledge (TSPK) in uncovering their incoming students' misconceptions then selecting appropriate instructional strategies, learning resources and content organisation to address these. TSPK is canonical, and is an important dimension of teacher professional knowledge and skill (Gess-Newsome, 2015). Teachers' practice also often echoes the common body of knowledge that is evidenced through research as they informally gather evidence of student knowledge and understanding. For example, it is known that students develop their misunderstandings through misinterpretation of observable everyday phenomena, despite explanations through theory taught in high school and undergraduate chemistry (Johnstone, 1983). Furthermore, it is known that students apply teleological reasoning to explain observed phenomena based on their understanding of the real world and analogies (Talanquer, 2007; Talanquer, 2013). These alternate conceptions extend into higher instructional and developmental levels; indeed many incoming chemistry doctoral students also find it difficult to apply their theoretical understanding of chemistry concepts to answer questions that have been situated in real-world contexts (Bodner, 1991). Over the past three decades, chemistry education researchers have built a strong collective knowledge surrounding alternative conceptions, including insights into the origin of misconceptions (Bergquist & Heikkinen, 1990; Hackling & Garnett, 1985; Hand & Treagust, 1988; Kind, 2004; Peterson, Treagust, & Garnett, 1989).

Conceptual change as a field of constructivist research has been the focus for widespread cognitive research studies in science education over many decades. Several reviews of the field (Driver & Erickson, 1983; Duit & Treagust, 2003; Duit & Treagust, 2012) describe the journey from the classical approach to conceptual change, deriving from the seminal work of Posner, Strike, Hewson, and Gertzog (1982), to current perspectives. Two approaches to thinking about conceptual change are popular; these are based on the work of diSessa (diSessa, 1988; diSessa, 2008; diSessa, Gillespie, & Esterly, 2004), who regards the process of learning as collecting pieces of knowledge and integrating them into a larger system of complex knowledge, and the work of Vosniadou (Vosniadou, 2012; Vosniadou, Vamvakoussi, & Skopeliti, 2008), who considers preconceptions to be ideas encountered informally that are built upon or require adjustment in formal learning. Recommendations from this body of research almost universally encourage further research into individual student conceptions with a goal of improving teachers' pedagogies and practices. This aligns with Ausubel's view (1968, p. vi) that 'If we had to reduce all of educational psychology to just one principle, we would say this: The most important single factor influencing learning is what the learner already knows. Ascertain this and teach him [*sic*] accordingly'. Conceptual change research studies have been disseminated across the gamut of science discipline-based education research journals, resulting in a vast array of strategies and understanding (Singer, Nielsen, & Schweingruber, 2012). In chemistry, the investigation of students' existing conceptions is a common pursuit, and research targeting specific concepts has expanded during the past decade. Singer and colleagues (2012, p. 60) observe that 120 chemistry research papers were published on the topic of conceptual change between 2000 and 2010. One particular approach adopted by researchers is to design multiple-choice items and assemble them into inventories (diagnostic instruments) to gain insight into student thinking.

Nearly 30 years ago, Treagust proposed a methodology for chemistry educators and researchers to translate knowledge of student misconceptions gained through qualitative observations and interviews into diagnostic instruments (Treagust, 1988). This methodology forms a widely accepted basis for the development of chemistry diagnostic instruments, and has been extended to formative assessment measuring effectiveness of instruction (Adams & Wieman, 2011). Chemistry conceptual change research has resulted in numerous diagnostic instruments that have been reported to explore a range of conceptions across multiple levels of learning. The instruments that have particular relevance to the present study are summarised in Table 1.

It can be seen that slightly over half of these instruments were developed for use with high school students and the remainder were developed for first-year undergraduate students; several have been used in both contexts. Of the instruments listed, the Chemical Concept Inventory (CCI) (Mulford & Robinson, 2002) is most widely adopted (Kruse & Roehrig, 2005) and has been recently validated through extensive psychometric testing (Barbera, 2013; Schwartz & Barbera, 2014). Concepts and topics that are addressed in the CCI include the conservation of matter, phase change, solutions and atomic properties. Approximately half of the items in the CCI were adapted from previous instruments and many were informed by published qualitative research studies.

The wide variety of concept inventories that are published annually addressing concepts across both science and engineering disciplines demonstrates the sustained interest in this form of assessment tool. However, the utility of the majority of these instruments, beyond their initial publication and use by their originating research group, is not commonly shared. An exception is the CCI; a review of subsequent use demonstrates that similar outcomes are obtained for many items regardless of context (Barbera, 2013; Kruse & Roehrig, 2005; Lawrie et al., in review; Schwartz & Barbera, 2014).

Multiple choice (MC) is the most popular format for diagnostic items, offering the advantage that large numbers of students can be tested and results can quickly be generated and distributed compared to written responses, interviews or open-ended surveys. The disadvantage of MC is that new insights into conceptual understanding cannot be obtained from student explanations unless they are invited to explain the reason for their choice. Two methods that have been developed to increase information into student understanding while maintaining the MC format are Ordered Multiple Choice (OMC) and two-tier (or multiple-tier) items.

OMC involves the careful design of distractors when developing MC diagnostic items to elicit awareness of the nature of students' conceptions. This is particularly effective when the distractors are based on known levels of conceptual development. OMC was established by Briggs and co-workers in 2006 (Briggs, Alonzo, Schwab, & Wilson, 2006) and has been adopted by Parchmann and colleagues for chemistry concepts (Hadenfeldt, Bernholt, Liu, Neumann, & Parchmann, 2013). Combining incorrect responses from several related items corrects for linguistic problems, poor concentration or simple transcription errors.

Two-tier items, pioneered by Treagust (1986), consist of an MC or a true/false question in the first tier coupled to a second question, in which students select the reason for their first response (Othman, Treagust, & Chandrasegaran, 2008; Peterson et al., 1989; Treagust, 1988; Tyson, Treagust, & Bucat, 1999). The second tier is developed, informed by student interviews and evidence of common misconceptions. Hence, like OMC, such items provide insights into student conceptual understanding. One limitation of two-tier items is the interpretation of instances when students select the incorrect answer in the first tier but the correct reason in the second tier. Despite this, two-tier instruments are popular and are used in combination with single-tier MC items in many instruments (Table 1).

Table 1. Chronological list of relevant published quantitative concept inventories, their composition and context.

Inventory	Concepts covered	Total number of items (of those, two tier) and format	Availability of complete inventory of items	Context	Reference
Covalent bonding and structure	Bond polarity, molecular shape, molecule polarity, lattices, intermolecular forces and octet rule	15 (15) MC	Not specified	High school (grade 11 and 12)	Peterson et al. (1989)
Untitled	Isomerism, redox and acid–base,	6 MC from large pool	Not specified	High school (grades 11–13)	Schmidt (1997)
Untitled	Equilibrium	Not specified; all two tier MC	Not specified	High school	Tyson et al. (1999)
Chemistry Concepts Inventory (CCI)	Conservation of mass-matter, phase change, stoichiometry/limiting reagent, bond energy, atomic scale, specific heat capacity and solutions	16 (6) MC	Supplementary material	First-year undergraduate	Mulford and Robinson (2002)
Chemistry I & Chemistry II	Specific heat capacity, bonding, phase changes, equilibrium, acids & bases and electrochemistry	30 + 31 MC	Contact authors	First-year undergraduate	Krause, Birk, Bauer, Jenkins, and Pavelich (2004); Pavelich, Jenkins, Birk, Bauer, and Krause (2004)
Chemistry concept test (CCT)	Stoichiometry, mole concept, intermolecular forces, conservation of matter, periodic properties, acids & bases, chemical equilibrium, electrochemistry and organic chemistry	65; used 37 and 37 with 9 common; 60 MC 5 SA	Not specified	First-year undergraduate	Potgieter and Davidowitz (2011); Potgieter, Davidowitz, and Blom (2005); Potgieter, Davidowitz, and Venter (2008)
ParNoMA	Particulate nature of matter	20 MC	Supplementary material	Middle school, high school and undergraduate	Yezierski and Birk (2006)
Two-concept diagnostic	Particulate nature of matter and bonding	10 (10) MC	Contact authors	High school (grade 9 and 10)	Othman et al. (2008)
General Chemistry I Concept Survey	Covalent bonding and molecular structure	26 MC	Not specified	First-year undergraduate	Pentecost and Langdon (2008)
Kinetic particle theory inventory (KPTI)	Interparticular spacing, changes in state, intermolecular forces, and diffusion in liquids and gases	11 MC	Supplementary material	High school (ages 14–16)	Treagust et al. (2010)
Structure and motion of matter SAMM	Particulate nature of matter	3 multipart open-ended	Supplementary material	Middle school, high school and undergraduate	Stains et al. (2011)
Untitled	Particulate nature of matter	10 OMC	Supplementary material	High school (grades 6–12)	Hadenfeldt et al., 2013
Thermochemistry Concept Inventory	Thermodynamics	10 MC	Contact authors	First-year undergraduate	Wren and Barbera (2014)
Bonding Representations Inventory	Bonding and representations of bonding	15 (8) MC	Contact authors	High school and undergraduate	Luxford and Bretz (2014)
Untitled	Conservation of mass-matter, phase change, stoichiometry/limiting reagent, bond energy, atomic scale and solutions	22 MC	Supplementary material	First-year undergraduate	Lawrie, Schultz and Wright (in review)

Thus, a substantial pool of items exists upon which teachers can draw to explore their own students' conceptions in their practice. However, the use of an existing instrument for diagnosing conceptions may be difficult because the specific set of topics that a teacher seeks to explore may not match any currently available instrument. Furthermore, most of the instruments in Table 1 were designed for educational research purposes and not for classroom diagnostic practices. Indeed, tension is emerging in the different ways that teachers and educational researchers apply these diagnostic tools and how they use the information gleaned. Some researchers aim to protect the integrity of their instruments by preventing widespread dissemination of their items, preferring to be contacted directly by users to provide access to items, while others openly disseminate their instruments to inform teaching practices and enhance teachers' TSPK (Stains, Escriu-Sune, De Santizo, Molina Alvarez de Santizo, & Sevian, 2011) or to examine how teachers apply diagnostic tools to gain insight into their students' thinking (Sadler, 1998).

Increased rigour has been called for to establish validity and reliability of research instruments (Arjoon, Xu, & Lewis, 2013; Barbera, 2013; Schwartz & Barbera, 2014; Wren & Barbera, 2013) and recently published instruments that explore conceptual understanding in a single topic have manifested exacting standards (Brandriet & Bretz, 2014; Luxford & Bretz, 2014; Wren & Barbera, 2013; Wren & Barbera, 2014). Teachers may often be unaware of the underlying process and intention that was involved in the development of these carefully structured, validated and reliable diagnostic instruments. As practitioners, they do not have the resources or time to conduct detailed student interviews to support the development of their own diagnostic items. They are also not equipped to validate instruments for each particular cohort, and are more likely to use items that appeal to them in combinations that suit their immediate purposes. Outcomes of ad hoc use of diagnostic items in practice are rarely disseminated in the research literature, so there is likely an extensive body of data in existence that is not published.

With these factors in mind, we have developed the following research questions aimed at exploring the process of developing a diagnostic instrument for the purposes of teaching practice and providing formative feedback to tertiary students.

- To what extent do concept items retain their validity when they are separated from the instrument in which they originated?
- To what extent does the structure and format of diagnostic items impact upon students' responses?

The context of this study was first-semester undergraduate general chemistry classes at five large research-intensive Australian universities located in three Australian states (Institutions 1 & 2 in Queensland, Institutions 3 & 4 in New South Wales and Institution 5 in Victoria). All these classes are large, with between 250 and 1500 students enrolled depending on the semester and institution; cohorts typically include between 10% and 20% international students. High school chemistry is a pre-requisite for entry into first-semester chemistry at only one of these institutions. Thus, the combined cohort of students entering tertiary chemistry typically represents diverse prior learning experiences and understandings in chemistry.

To date no published instrument exists that is designed to provide strategic and constructive formative feedback to students about their conceptions. In this study, an instrument with this aim has been validated. The provision of formative feedback is recognised as a transitional pedagogy (McInnis, James, & McNaught, 1995; Yorke, 2001) aligned with high-impact assessment and feedback practices (Nicol & Macfarlane-Dick, 2006). This manuscript reports part of a larger study (Lawrie et al., 2013; Lawrie et al., 2015) and forms the continuation of our previous work developing diagnostic instruments for practitioners (Lawrie et al., in review). The overarching aim was to partner tailored formative feedback to students with associated remediation activities (Lawrie et al., 2016).

Methodology

The project team, 10 faculty members who identified as chemistry education researchers, had substantial knowledge of the extensive published research in exploring student existing preconceptions and alternative conceptions in chemistry. As the first step in the process of item selection and development, five core topic areas were selected to align with topics taught in the general chemistry courses at all five institutions: phase change, heat and energy, conservation of matter, aqueous solutions and equilibrium. This topic choice was informed by research literature on concepts that present persistent difficulties for students (Johnstone, 2010). The topics also aligned with five of the big ideas of undergraduate chemistry that have been elucidated following an exhaustive process choreographed by the American Chemical Society's Examinations Institute (Murphy, Holme, Zenisky, Caruthers, & Knaus, 2012). For each of these topics, we were aware of validated items that we could use or adapt for our instrument.

A process was established for the identification and refinement of items involving collation and evaluation of published instruments on each topic to form clusters of five questions for each topic. Where necessary to complete a cluster, new items were developed according to the methodology first proposed by Treagust (1988). Details of this process together with perceptions of our faculty community of practice will be published separately (Lawrie et al., in preparation). The principles behind the OMC strategy represented a viable mechanism for the provision of richer formative feedback to students than simply informing them whether their answer was correct. This strategy enabled combination of their incorrect responses within a specific topic cluster to highlight specific misconceptions or ill-developed conceptions. As part of the process of adopting this approach, items that had been sourced from existing instruments, which were not developed as OMC items, were re-appraised by the project team to assign their distractors to different developmental levels (Vosniadou, 2012), as described below for one cluster.

While considering the distractors for items in the diagnostic tool, the project team deliberated extensively on the best approach to encourage students to answer authentically so that effective formative feedback could be delivered. It was decided that, where feasible, items would include a distractor 'I don't know', to avoid 'forcing' students to guess and then select one of the options randomly, which would result in delivery of ineffective feedback. Since our cohorts include students who have never studied chemistry before, there will be some students who genuinely cannot answer some items. Our intention was to encourage students to engage with the instrument so that we could help remediate their missing or incorrect conceptions with tailored feedback. Also, if students are intimidated by their lack of prior knowledge and feel unprepared when they encounter terms that they have not heard, it is possible that some students may withdraw. The literature on including 'I don't know' as an answer option does not present a consistent view (Haladyna &

Downing, 1989; Haladyna, Downing, & Rodriguez, 2002). Note that the proportion of students answering 'I don't know' was fewer than 2% for all items of our instrument.

After preliminary assembly of the instrument, all items were re-assessed to consider content and face validity along with reliability (Wren & Barbera, 2013). A number of existing items required minor amendment to phrasing and terminology for the Australian context while others were reworded to better align them to the topic cluster. A consistent format and presentation in the display of items was adopted including modification of any sub-microscopic representations, applying best practice in diagrammatic representations of chemical structures (Tasker, 2014). The origin of all items that were used (Table Supp1) and the changes made to each between 2013 and 2014 (Table Supp2) are provided in the supplemental material together with the complete final instrument (Table Supp2).

Phase change cluster

Understanding of phase transformations is built upon the concepts of the particulate nature of matter, intermolecular forces and fundamental ideas related to atoms and molecules. Significant knowledge regarding students' understanding of phase change exists (Bridle & Yezierski, 2012; Coştu, Ayas, & Niaz, 2010; Johnson, 1998; Kirbulut & Beeth, 2013; Mulford & Robinson, 2002; Othman et al., 2008; Yezierski & Birk, 2006). The project team decided that water should be the common substance for all items in this cluster because the properties of water have been the focus of substantial educational research, are used in instructional resources for related topics in general chemistry and also have relevance to the real world. For example, item 4 in this cluster was derived from the CCI (Mulford & Robinson, 2002) where the context was a glass of cold milk. This was adapted from the original qualitative work published by Osborne and Cosgrove in 1983 involving ice. We decided to return to the original focus of the item, that is, to water for consistency in our item cluster. The cluster was thus designed to gain insights into student understanding of the processes of boiling, evaporation and condensation of water, and enable provision of relevant formative feedback.

It has been reported that the word 'water' is sometimes used by learners to mean any liquid (Krnel, Watson, & Glažar, 2005; Solomonidou & Stavridou, 2000), which may imply that water has an additional dimension when used to explore concepts and therefore could represent a weakness in our decision to use only water as the context in the cluster. However, those studies involved younger learners who were developing their language in parallel with their conceptual understanding. It is also established that students' development of their conceptions changes over time because they experience ontological shifts in their thinking (Tytler, 2000). In our study, it was assumed that first-year students represent adult learners who possess more sophisticated ontologies and epistemologies; hence our choice of water as the common molecule should not be problematic.

Applying the underlying principle of OMC, the distractors for the five items within the phase change cluster were classified into four categories (Figure 1). The categories chosen for developmental stages of this concept are physical change (correct response), chemical change, teleological responses and 'I don't know'. These categories were informed by the body of research regarding students' thinking around phase changes involving water and also by published outcomes from the CCI (Barbera, 2013; Mulford & Robinson, 2002; Schwartz & Barbera, 2014). The most common alternate conception is that water undergoes

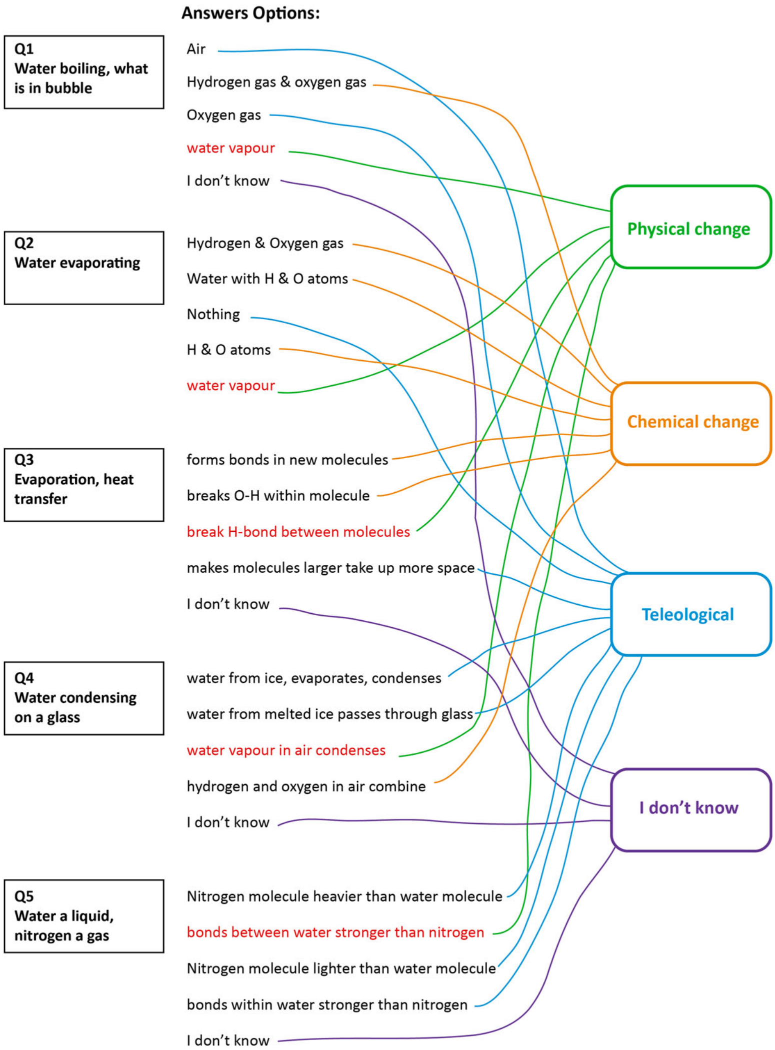

Figure 1. Classification of distractors into categories that describe different levels of student understanding for the phase change cluster.

a chemical change into hydrogen and oxygen molecules or atoms. Teleological explanations were distractors that involved a purpose or a consequence (Talanquer, 2007; Talanquer, 2013). For example, air and oxygen gases are known to be dissolved in water in the context of streams or fish tanks. Figure 1 illustrates the relationships between the distractors and the categories, indicating how student responses to the items within the cluster can be integrated to provide formative feedback based on the student's level of understanding. It can be seen that, while these items were not originally designed to function as OMC items, this strategy could be retrospectively applied to the items in a meaningful way.

Heat and energy cluster

The topic of heat and energy is complex and is strongly related to learning progressions associated with energy extending from primary school to tertiary education (Lewis &

Linn, 1994; Prince, Vigeant, & Nottis, 2012). This is a critical topic in general chemistry that, while considered widely in physics and engineering education research, has received less attention in chemistry concept research. This may be due to the crosscutting nature of the concepts that underpin bond energies, heat transfer, temperature, phase change and equilibrium (Chu, Treagust, Yeo, & Zadnik, 2012). Items for this cluster were selected or developed by the project team with the objective of providing students with feedback specifically in regard to their understanding of the flow or transfer of thermal energy and the breaking and formation of bonds.

Conservation of matter cluster

This cluster was designed to explore student understanding of the particulate nature of matter in chemical reactions. We selected this topic to build on the outcomes of our earlier study (Lawrie et al., in review) where students found it difficult to account for atoms in a chemical reaction. The understanding of concepts of matter has been widely researched, with many studies focusing on how students construct and apply their understanding from early conceptions and misconceptions (Kahveci, 2013; Stains et al., 2011; Talanquer, 2016; Treagust et al., 2011).

An example of a modification that was made to an item to fit a cluster is for item Q15 in this cluster, which had the stem 'when a match burns, matter is destroyed'. This was originally a true/false question representing the first tier of a two-tier item in the CCI. Published outcomes for this item indicated that it was too easy and non-discriminatory (Barbera, 2013; Mulford & Robinson, 2002; Schwartz & Barbera, 2014). Thus, it was reformulated into a single-tier item by embedding the reasoning from the second tier to form the distractors to provide more effective formative feedback.

Aqueous solution cluster

Student understanding of speciation in aqueous solutions has been widely researched (Devetak, Vogrinc, & Glažar, 2009; Dickson, Thompson, & O'Toole, 2016; Kahveci, 2013). Concepts related to this topic have attracted interest from researchers for a number of reasons: insight is gained into students understanding of ionic bonding (Othman et al., 2008), speciation is important for explaining experimental observations such as precipitation (Tasker, 2014) and students struggle with quantitative calculations relating to solution concentration and dilution (de Berg, 2011; Devetak et al., 2009; Jansoon, Coll, & Somsook, 2009). It is well established that many students do not possess a mental model of separated, hydrated ions in solution and this impacts on their ability to apply simple ratios in dilution calculations. The focus of this cluster was on the identity of dissolved species and the interpretation of concentration. Very few items had been reported in the literature that suited the aim for this cluster and hence four items were developed by the project team, informed by their prior research (Dickson et al., 2016; Tasker, 2014; Tasker & Dalton, 2006).

Equilibrium cluster

Equilibrium has long been regarded as one of the most problematic areas for chemistry students because it relies on several fundamental concepts that students must integrate

(Bergquist & Heikkinen, 1990; Hackling & Garnett, 1985; Tyson et al., 1999). The inherent scientific terminology, including words that have different meanings in everyday life, makes this topic area particularly challenging (Johnstone, 2010). This topic was chosen to address higher level conceptions typically encountered in high school chemistry that are extended in the tertiary context. Since equilibrium concepts are central to understanding multiple topics (e.g. solutions, acids & bases, phase transitions), it proved difficult to collate existing literature items into a coherent equilibrium cluster so several items were developed by the team. These items addressed concepts relating to Le Châtelier's principle, dissolution, saturated solutions and graphical representations of concentration changes.

In 2013, the resulting 25-item instrument was delivered at the beginning of the first semester of first-year chemistry at four of the institutions. While high school chemistry preparation was recommended for enrolment at all five institutions, only Institution 1 required students to have this preparation; at the other institutions, between 10% and 30% of students had not studied chemistry before. All enrolled students were invited to complete an online questionnaire (Qualtrics^{TM}) by email in week 1 of semester 1. Participation in this research study was on a voluntary opt-in basis through informed consent and institutional ethical approval was gained. Demographic questions embedded in the questionnaire enabled comparison of outcomes for students for whom English was a second language (ESL).

Data were filtered for consent and completion, then combined and statistically analysed (SPSS). Descriptive statistics included the mean and standard deviation values for each item to obtain the % correct responses. Classical test theory was applied to determine item difficulty, discrimination, effect size (Hedge's *g* variation of Cohen's *d* was applied to account for different population sizes where applicable) and Cronbach's alpha values (for each cluster of five items as well as the whole instrument). The discrimination for each item was determined through point biserial analysis regarded as most suitable for dichotomous data. Data were also filtered to compare the outcomes for students who had completed high school chemistry preparation with those who had not. Rasch analysis was also completed to evaluate the instrument's internal validity (Rasch, 1993).

While this instrument was intended ultimately to be a teaching practitioner's tool, one aim of this study was also to consider whether the items sourced from existing research-based diagnostic instruments retained their cited ability to discriminate students' conceptions. Therefore, evaluation of the whole instrument and individual items has been completed. Data were combined from all the participating institutions.

Results & discussion

To enable the provision of timely formative feedback to students, the instrument was delivered online during the first two weeks of the first semester across four participating institutions in three Australian states in 2013. A total of 1988 students attempted and 1132 students completed the instrument out of a possible 3517 enrolled students at the four institutions. Three institutions included students within their cohorts who had not completed high school chemistry despite advice that this preparation was required for success. Therefore, it was important to explore whether completion of high school chemistry impacted substantially on student outcomes. These are presented for each item, filtered by prior chemistry experience, in Figure 2. Significant differences were observed

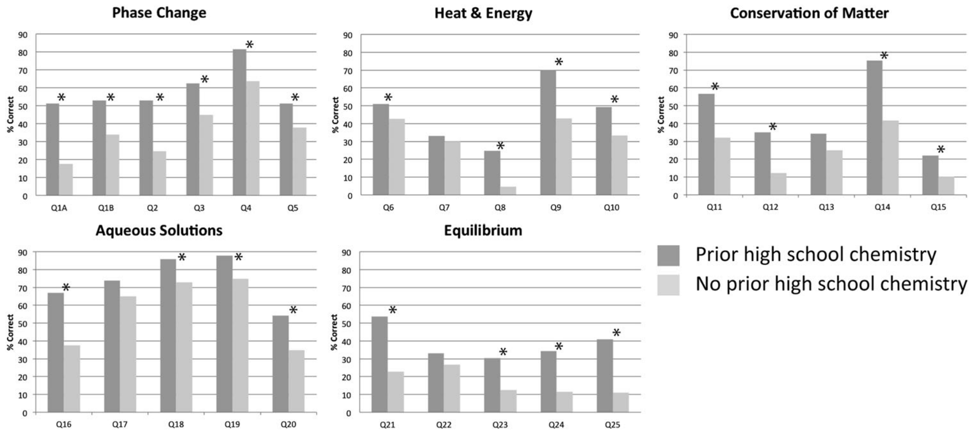

Figure 2. Percentage correct for each item in 2013, illustrating the impact of high school chemistry ($N = 1089$ prior high school chemistry, $N = 155$ no prior high school chemistry). Asterisks indicate significant differences.

between students who had completed high school chemistry compared to those who had not for every item except Q7 and Q22, with medium to large effect sizes observed (*t* statistics and effect sizes are provided in the supplemental material Table Supp3).

Classical test theory includes the determination of item difficulty (mean score for each item) and evaluation of whether each item is able to discriminate between different levels of student understanding (point biserial analysis). Seventeen items were found to lie within the accepted ranges for difficulty (0.25–0.75) and discrimination (equal to or above 0.3), highlighted by the box shown in Figure 3. Eight items were identified as problematic (numbered outside the box in Figure 3) including six that were insufficiently discriminatory. One item was considered too difficult based on this analysis (Q15 in the conservation of matter cluster) and a second item, while discriminatory, was deemed insufficiently difficult (Q4 in the phase change cluster).

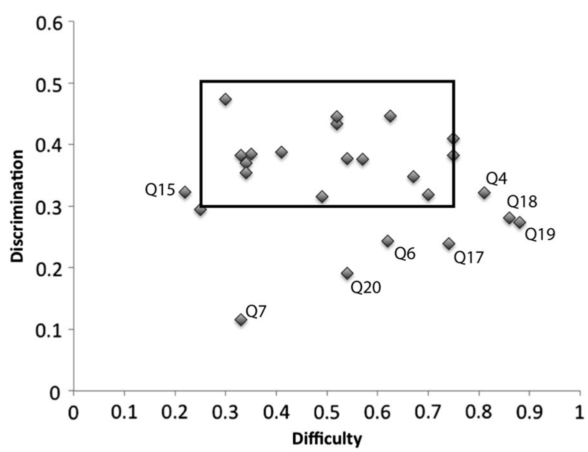

Figure 3. Comparison of discrimination and difficulty for students who had completed high school chemistry. Items that are in the accepted range of values for a concept inventory are contained within the box. Data are from 2013.

While several items were identified as non-discriminatory ($D < 0.3$), they were retained in their clusters because they complemented other items within the same cluster enabling OMC diagnosis of alternate conceptions for individual students. Q18 and Q19 were evidently the simplest items (greater than 75% students selected the correct answer); so they were substituted with newly developed items in the 2014 delivery of the diagnostic instrument.

A parallel measure of item difficulty is the logit scale derived from Rasch analysis, used in previous related studies of the CCI items (Barbera, 2013, Lawrie et al., in review). Published logit values for individual items have been compared with the values obtained in this study (Table Supp4) and general agreement was observed, with major differences only when an item was changed from two tier to single tier.

Rasch analysis of the whole instrument resulted in a Wright map representing the person and item internal consistency (Cronbach's alpha value of 0.81) and this has been displayed to show items in their clusters (Figure 4). It is evident that the equilibrium cluster contained

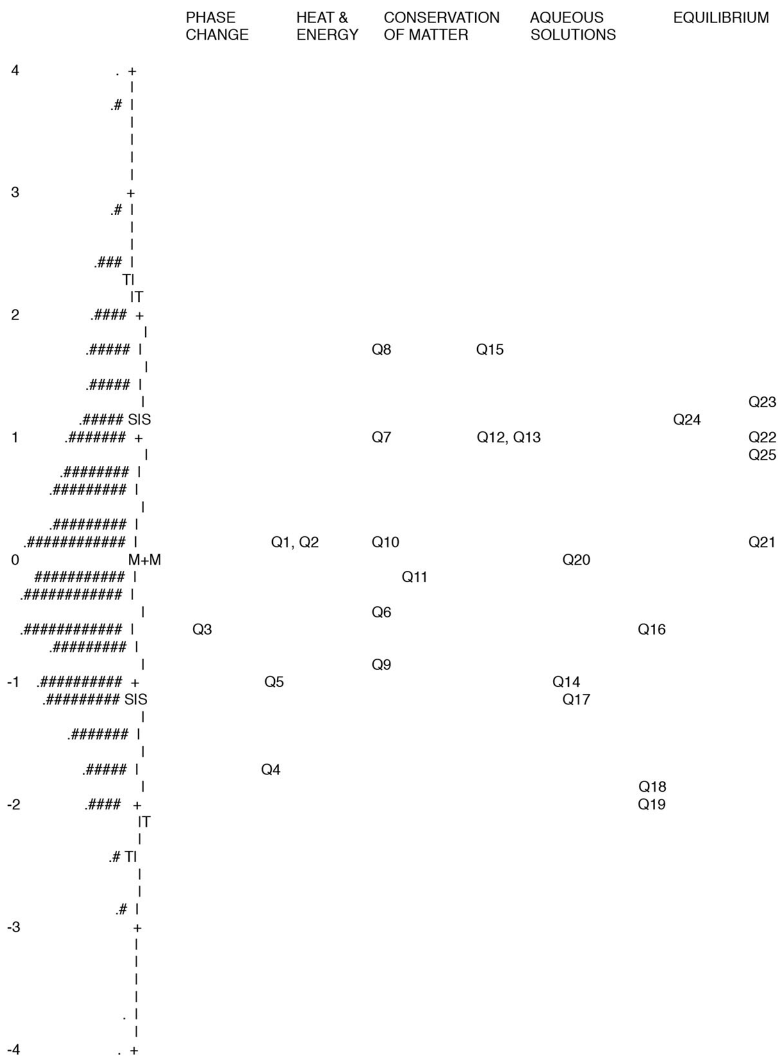

Figure 4. Wright map generated from Rasch analysis of the 2013 data (# represents eight participants).

the most difficult conceptual items with all five items (Q21–25) displayed as positive logit values. The aqueous solutions cluster has all five items (Q16–20) with negative logit values indicating that this is the easiest cluster. This correlates with the data shown in Figure 3 where Q17–20 were highlighted as non-discriminatory and not very difficult.

Having validated the instrument and established that items drawn from literature and new items could be combined to form a single instrument, several variables that might impact on diagnosis of students' prior understanding of chemistry were considered more deeply. Linguistic (Johnstone, 1991; Taber, 2015) and cultural influences (Solano-Flores & Nelson-Barber, 2001) on item validity are strongly dependent on the context in which they are placed. In Australia, there is a significant cohort of tertiary students for whom English is a second language (ESL); so it was critical to try to ensure that incorrect responses were not attributable to an issue of vocabulary or unfamiliar language usage. A comparison of outcomes for students with English as their first language and ESL students revealed significant differences for 11 items and these data are presented in Table 2.

Of the 11 items for which a significant difference was observed, the ESL students performed better in 7 items. Many international students have completed up to 12 months more of high school chemistry compared to Australian high school leavers; so it is inferred that their additional experience of chemistry led to a higher level of conceptual understanding compared to Australian domestic students. Of the four items in which ESL students performed worse than students who had self-identified as having English as their first language, each possessed a potential barrier in the form of terminology either in the item stem or the distractors. Vocabulary or terms that ESL students may not be familiar with included: 'railway track' (Q9); 'fate' and 'ash' (Q15); 'dilution' (Q17) and 'evaporation' (Q18). However, none of the statistical effect sizes is large for this variable; the effect sizes are small for eight of the items, and small to medium for Q7, Q8 and Q17.

The above combination of psychometric analyses based on classical test theory confirmed that the instrument functioned as a valid diagnostic tool. However, several items

Table 2. Statistical comparison of the outcomes for the core concept instrument between students who self-identified as ESL and the remainder of the cohort ($N = 1087$).

Concept item	Topic	English as first language M (SD)	ESL M (SD)	t statistic	Significance	Effect size (Cohen's d*)
Q5	Intermolecular forces (nitrogen gas and water)	0.74 (0.439)	0.82 (0.391)	−2.194	.029	0.17
Q6	Definition of heat	0.60 (0.491)	0.74 (0.442)	−3.752	<0.001	0.29
Q7	Energy transfer (bond formation)	0.30 (0.459)	0.49 (0.501)	−4.582	<.001	0.41
Q8	Bond energies (dinitrogen tetroxide)	0.22 (0.413)	0.40 (0.491)	−4.515	<.001	0.42
Q9	Heat effect (iron railway expansion)	**0.71 (0.452)**	**0.63 (0.485)**	2.212	.028	0.19
Q10	Heat and temperature (plastic and metal)	0.47 (0.499)	0.59 (0.493)	−2.837	.005	0.24
Q14	Stoichiometry	0.74 (0.439)	0.83 (0.377)	−2.807	.005	0.21
Q15	Conservation (fate) of matter	**0.23 (0.421)**	**0.16 (0.371)**	2.122	.035	0.16
Q16	Dissolution (calcium chloride)	0.65 (0.477)	0.76 (0.425)	−3.157	.002	0.24
Q17	Dilution (sugar solution)	**0.76 (0.424)**	**0.60 (0.492)**	4.176	<.001	0.39
Q18	Salt solution evaporation	**0.87 (0.333)**	**0.78 (0.417)**	2.836	.005	0.28

*Values are Hedge's g variation of Cohen's d taking into account the two different population sizes.

were identified as requiring modification or substitution with the aim of developing a more discriminatory and challenging suite of concept clusters. The project team also decided that only single-tier OMC items would be used to improve the useful formative feedback for students. Each item in the diagnostic instrument was considered in terms of student outcomes, validity and reliability, difficulty and discrimination, as well as revisiting alternative conceptions that students selected. Details of the revision, reformatting or substitution of items have been provided in the supplemental material (Table Supp5).

In particular, during this review process, the project team became aware of the role of visual representation as potentially influencing how students answered an item (Cook, Wiebe, & Carter, 2008; Crisp & Sweiry, 2006). It was noted that Q1 had been presented in published studies with (Johnson, 1998; Osborne & Cosgrove, 1983; Othman et al., 2008) and without (Mulford & Robinson, 2002) a supporting visualisation image (not integrated into the item); hence a question arose in regard to whether this representation provided scaffolding that supported students in their understanding. To explore this further, the team developed a graphical representation of a beaker with bubbles. In 2013, all participating students were randomly assigned either a text only (Q1A) or a graphical version (Q1B) of the same item when they accessed their online questionnaire. It was found that there was no significant difference in the outcomes for Q1A and Q1B for students who had completed high school chemistry. However, there was a significantly higher achievement by students who had no prior high school chemistry who had been presented with the graphical version of the item (full data in Table Supp3). This indicated that visual representations should be included in the selection of resources to scaffold understanding. Use of diagrams and schematic representations within diagnostic items should follow the best practice in chemical visualisation to avoid generating misunderstanding (Tasker, 2014). Four other items (Q2, Q8, Q12 and Q17) contained a graphical display of the molecular level information in the core instrument but there was no pattern of higher achievement evident for these items compared with the text-only items.

In 2014, the primary goal for the project team was to deliver effective formative feedback to their own students; hence the delivery in terms of the timing and structure of the instrument was highly variable. The diagnostic items were applied in separate concept clusters rather than a single delivery of the entire instrument at three institutions, and timing aligned with corresponding teaching activities and sequence of topics according to the context and curriculum at each institution. Figure 5 illustrates the process and timing; note that only two of the institutions in the study used all five clusters in 2014.

The student outcomes for individual items for Institution 1 was compared between 2013 and 2014, and data are presented in Figure 6. It is evident that there was a consistent performance for items used in both these years. Improvements in performance in 2014 may be attributed in part to the timing of their delivery, with clusters of questions used in parallel with the teaching schedule rather than all presented in week 1 as in 2013. It can also be seen that students found some topics less challenging (phase change and aqueous solutions) than others (conservation of matter and equilibrium). A complete statistical analysis for the items used in both years is provided in the supplemental material (Table Supp6).

It is evident that the approach of applying an OMC strategy to categorising distractors provides rich feedback to the instructor in regard to the range of conceptions held by their students but also in regard to troublesome concepts where the majority of the class opted

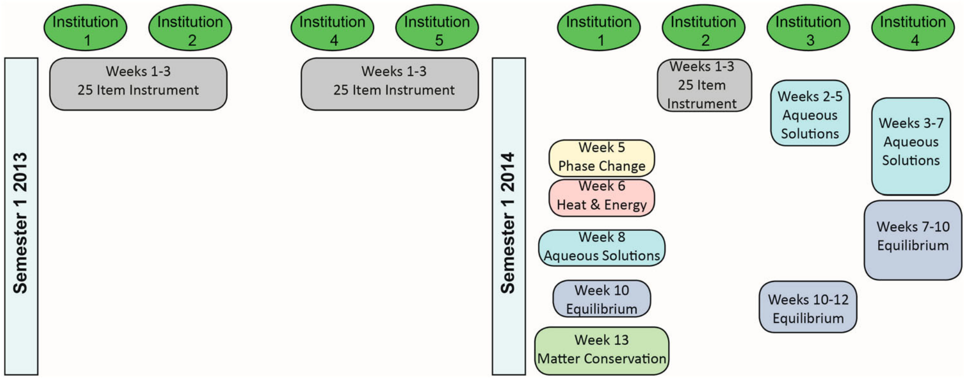

Figure 5. Timeline for the delivery of the concept clusters in 2014 illustrating the different applications at each institution.

for one incorrect response. Of particular interest is the conservation of matter (Q15) where less than 30% of students selected the correct answer in both 2013 and 2014 (Institution 1), with the incorrect responses indicating that their misunderstanding of the concept implied that heat possessed mass. The next step for the instructor is to use this information to embed a learning activity that addresses students' alternate conceptions, either as part of the face-to-face teaching sessions or as a self-directed online learning module (Lawrie et al., 2016, Lawrie et al., in review, Lawrie et al., 2015).

One further objective of this study was to explore the potential limitations of statistical analyses that are commonly applied to concept inventory data relating to the number of items considered. For comparison with 2014 data, the 2013 data were subdivided into topic clusters and the discrimination index (point biserial analysis) was calculated for subsets of five items as well as the full instrument (supplemental material Table Supp7). The resulting data are difficult to interpret; it was observed that, when considered as

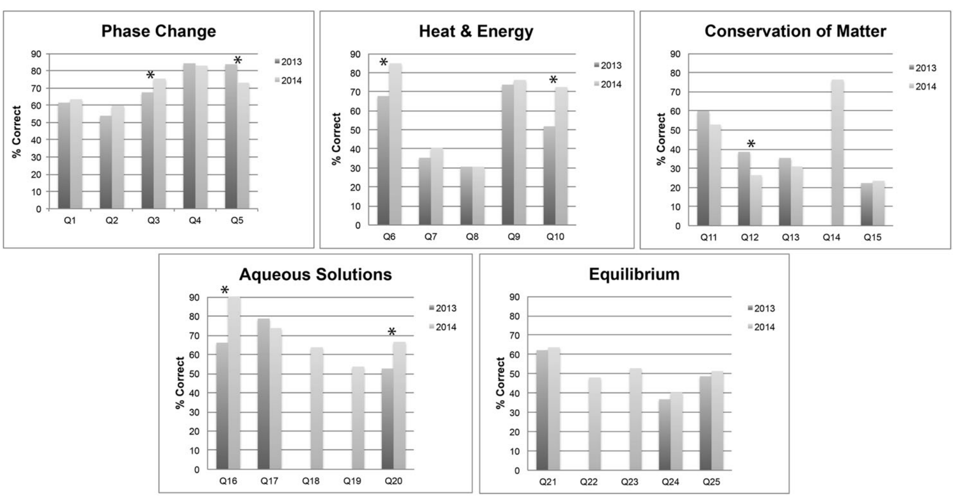

Figure 6. Percentage of students correct for 2014 items; data for 2013 are included for items that remained the same between 2013 and 2014. Asterisks indicate significant differences.

Table 3. Cronbach's alpha values for each topic cluster. The 2013 data were collected at a single time point at the commencement of the semester and the 2014 data were collected at several time points during the semester.

	Phase change	Heat & energy	Conservation of matter	Aqueous solutions	Equilibrium	Combined clusters
2013	0.63	0.44	0.55	0.45	0.62	0.81
2014	0.66	0.37	0.33	0.15	0.44	0.72

subsets of data in clusters, the discrimination of many of individual items dropped below the acceptable limit of 0.3. In contrast, the discrimination of several items increased when considered as a subset of data within their topic cluster (Q1, Q2, Q7, Q8, Q18, Q19, Q22 and Q23).

To further illustrate challenges in the statistical interpretation of data, the Cronbach's alpha values were calculated for each separate topic cluster and compared to the entire 25-item instrument. These data are presented in Table 3. The reader is reminded that the 2014 data were collected at five separate time points during the semester; hence it is questionable whether the alpha value of 0.72 for the combined data is valid. Cronbach's alpha values are known to be sensitive to both the number of items and the number of participants (Gliem & Gliem, 2003; Sijtsma, 2009; Tavakol & Dennick, 2011) and hence the low values for each cluster reflect the subdivision into five items. This effect is illustrated by the combination of the responses for the aqueous solutions and equilibrium clusters (10 items) for all participants across four institutions – the calculated Cronbach's alpha rises to 0.47 (2014) in comparison to the separate cluster values of 0.15 and 0.44.

While acknowledging the limitations of a small-item pool, the Cronbach's alpha for the phase change topic is higher than that for each of the other clusters (just below the commonly cited acceptable threshold of 0.7 for internal reliability). It is of interest to note that the items within this cluster possessed the greatest degree of contextual integration in terms of linked conceptions based on the choice of water as a common substance. It appears that this cluster functioned well in increasing insight into student thinking and enables the provision of useful feedback relating to alternate or missing conceptions. As was shown in Figure 1, each alternate conception was probed through several distractors, and so their combinations of responses can be used to engage students in a suitable learning intervention. The poor internal consistency for the other separate clusters reinforces the tension between practice and research in the application of concept diagnostic tools. For classroom practice, where the intention is to provide feedback, this may not represent a hurdle; however, the outcome of this study indicates that further exploration is required to develop small-item clusters for the purpose of researching the nature of alternate conceptions.

Summary and recommendations for practice

We have developed a useful diagnostic instrument organised into five topic clusters that can be delivered to support the timing and sequence of typical general chemistry courses. The outcomes of this diagnostic instrument provided us with deeper insight into the prior understanding held by incoming tertiary chemistry students enabling provision of formative feedback. We have included our complete instrument in the supplemental material and encourage its adoption in teaching practice.

Our data support the premise that it is possible to combine items sourced from multiple instruments to suit a range of teaching contexts. As a research tool, psychometric analysis of our data indicated that, when considered as clusters, subsets of five items demonstrated reduced validity. However, the validity is sufficient for teaching practice where it is not as important that an item be demonstrated as valid in each context of its use. Thus, items sourced from validated and reliable instruments can be combined as needed; however, the new instruments will require some degree of validation for the purpose of publication (Adams & Wieman, 2011).

In this study, we established that use of a pictorial representation to support a text item assisted students with a weaker background in chemistry to identify the correct response. In the discipline of chemistry where it has long been known that some students struggle with the macroscopic–submicroscopic–symbolic triad (Johnstone, 1991), supporting students by providing another representation is recommended, especially in the context of a diagnostic instrument being used to support learning rather than in a summative manner. Finally, we recommend using simple yet precise language in concept items and avoiding terms that are not necessary to test the concept under investigation and may confuse students who face language barriers.

Acknowledgements

The views expressed in this activity do not necessarily reflect the views of the Australian Government Office for Learning and Teaching. The significant contributions of project team members Hayden Dickson, Aaron Micallef and Md Abdullah Al Mamun to the development of tools and resources is gratefully acknowledged. We thank the students involved in the study. We also acknowledge insights from exchanges and discussions with the authors (Lawrie) regarding the application, analysis, validation of inventories including access to their own concept inventories by Jazilah Othman, Stephen Krause, Jennifer Lewis, Thomas Pentecost and Stacey Lowery-Bretz.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Support for this activity has been provided by the Australian Government Office for Learning and Teaching (ID12-2277).

References

Adams, W. K., & Wieman, C. E. (2011). Development and validation of instruments to measure learning of expert-like thinking. *International Journal of Science Education*, *33*, 1289–1312. doi:10.1080/09500693.2010.512369

Arjoon, J. A., Xu, X., & Lewis, J. E. (2013). Understanding the state of the art for measurement in chemistry education research: Examining the psychometric evidence. *Journal of Chemical Education*, *90*, 536–545. doi:10.1021/ed3002013

Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. New York, NY: Holt, Rinehart & Winston.

Barbera, J. (2013). A psychometric analysis of the chemical concepts inventory. *Journal of Chemical Education*, *90*, 546–553. doi:10.1021/ed3004353 de Berg, K. (2011). A study of first-year chemistry students' understanding of solution concentration at the tertiary level. *Chemistry Education Research and Practice*, *13*, 8–16. doi:10.1039/c1rp90056k

Bergquist, W., & Heikkinen, H. (1990). Student ideas regarding chemical equilibrium: What written test answers do not reveal. *Journal of Chemical Education*, *67*, 1000–1003. doi:10.1021/ed067p1000

Biggs, J. (1993). What do inventories of students's learning processes really measure? A theoretical review and clarification. *British Journal of Educational Psychology*, *63*, 3–19. doi:10.1111/j.2044-8279.1993.tb01038.x

Bodner, G. M. (1991). I have found you an argument: The conceptual knowledge of beginning chemistry graduate students. *Journal of Chemical Education*, *68*, 385–388. doi:10.1021/ed068p385

Brandriet, A. R., & Bretz, S. L. (2014). The development of the redox concept inventory as a measure of students' symbolic and particulate redox understandings and confidence. *Journal of Chemical Education*, *91*, 1132–1144. doi:10.1021/ed500051n

Bridle, C. A., & Yezierski, E. J. (2012). Evidence for the effectiveness of inquiry-based, particulate-level instruction on conceptions of the particulate nature of matter. *Journal of Chemical Education*, *89*, 192–198. doi:10.1021/ed100735u

Briggs, D. C., Alonzo, A. C., Schwab, C., & Wilson, M. (2006). Diagnostic assessment with ordered multiple-choice items. *Educational Assessment*, *11*, 33–63. doi:10.1207/s15326977ea1101_2

Chu, H.-E., Treagust, D. F., Yeo, S., & Zadnik, M. (2012). Evaluation of students' understanding of thermal concepts in everyday contexts. *International Journal of Science Education*, *34*, 1509–1534. doi:10.1080/09500693.2012.657714

Cook, M., Wiebe, E. N., & Carter, G. (2008). The influence of prior knowledge on viewing and interpreting graphics with macroscopic and molecular representations. *Science Education*, *92*, 848–867. doi:10.1002/sce.20262

Coştu, B., Ayas, A., & Niaz, M. (2010). Promoting conceptual change in first year students' understanding of evaporation. *Chemistry Education Research and Practice*, *11*, 5–16. doi:10.1039/C001041N

Crisp, V., & Sweiry, E. (2006). Can a picture ruin a thousand words? The effects of visual resources in exam questions. *Educational Research*, *48*, 139–154. doi:10.1080/00131880600732249

Devetak, I., Vogrinc, J., & Glažar, S. A. (2009). Assessing 16-year-old students' understanding of aqueous solution at submicroscopic level. *Research in Science Education*, *39*, 157–179. doi:10.1007/s11165-007-9077-2

Dickson, H., Thompson, C. D., & O'Toole, P. (2016). A picture is worth a thousand words: Investigating first year chemistry students' ability to visually express their understanding of chemistry concepts. *International Journal of Innovation in Science and Mathematics Education*, *24*, 12–23.

Driver, R., & Erickson, G. (1983). Theories-in-action: Some theoretical and empirical issues in the study of students' conceptual frameworks in science. *Studies in Science Education*, *10*, 37–60. doi:10.1080/03057268308559904

Duit, R., & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education*, *25*, 671–688. doi:10.1080/09500690305016

Duit, R., & Treagust, D. F. (2012). How can conceptual change contribute to theory and practice in science education?. In B. J. Fraser, K. G. Tobin, & C. J. McRobbie (Eds.), *Second international handbook of science education* (pp. 107–118). Dordrecht: Springer Science+Business Media.

Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK: Results of the thinking from the PCK summit. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 28–42). New York, NY: Routledge.

Gliem, J. S., & Gliem, R. R. (2003). *Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for likert-type scales*. Paper presented at the 2003 Midwest Research to Practice Conference in Adult, Continuing, and Community Education, Columbus, Ohio.

Hackling, M. W., & Garnett, P. J. (1985). Misconceptions of chemical equilibrium. *European Journal of Science Education*, *7*, 205–214.

Hadenfeldt, J. C., Bernholt, S., Liu, X., Neumann, K., & Parchmann, I. (2013). Using ordered multiple-choice items to assess students' understanding of the structure and composition of matter. *Journal of Chemical Education*, *90*, 1602–1608. doi:10.1021/ed3006192

Haladyna, T. M., & Downing, S. M. (1989). Validity of a taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*, *2*, 51–78.

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, *15*, 309–333. doi:10.1207/S15324818AME1503_5

Hand, B. M., & Treagust, D. F. (1988). Application of a conceptual conflict teaching strategy to enhance student learning of acids and bases. *Research in Science Education*, *18*, 53–63. doi:10.1007/BF02356580

Jansoon, N., Coll, R. K., & Somsook, E. (2009). Understanding mental models of dilution in Thai students. *International Journal of Environmental & Science Education*, *4*, 147–168.

Johnson, P. (1998). Children's understanding of changes of state involving the gas state, part 1: Boiling water and the particle theory. *International Journal of Science Education*, *20*, 567–583. doi:10.1080/0950069980200505

Johnstone, A. H. (1983). Chemical education research: Facts, findings, consequences. *Journal of Chemical Education*, *60*, 968–971. doi:10.1021/ed060p968

Johnstone, A. H. (1991). Why is science difficult to learn? Things are seldom what they seem. *Journal of Computer Assisted Learning*, *7*, 75–83. doi:10.1111/j.1365-2729.1991.tb00230.x

Johnstone, A. H. (2010). You can't get there from here. *Journal of Chemical Education*, *87*, 22–29. doi:10.1021/ed800026d

Kahveci, A. (2013). Diagnostic assessment of student understanding of the particulate nature of matter: Decades of research. In G. Tsaparlis, & H. Sevian (Eds.), *Concepts of matter in science education* (pp. 249–278). Dordrecht: Springer Science+Business Media.

Kind, V. (2004). *Beyond appearances: Students' misconceptions about basic chemical ideas* (2nd ed.). Royal Society of Chemistry: London.

Kirbulut, Z. D., & Beeth, M. E. (2013). Consistency of students' ideas across evaporation, condensation, and boiling. *Research in Science Education*, *43*, 209–232. doi:10.1007/s11165-011-9264-z

Krause, S., Birk, J., Bauer, R., Jenkins, B., & Pavelich, M. J. (2004). *Development, testing and application of a chemistry concept inventory*. Paper presented at the *34th ASEE/IEEE frontiers in education conference*, Savannah, Georgia, USA.

Krnel, D., Watson, R., & Glažar, S. A. (2005). The development of the concept of 'matter': A cross-age study of how children describe materials. *International Journal of Science Education*, *27*, 367–383. doi:10.1080/09500690412331314441

Kruse, R. A., & Roehrig, G. H. (2005). A comparison study: Assessing teachers' conceptions with the chemistry concepts inventory. *Journal of Chemical Education*, *82*, 1246–1250. doi:10.1021/ed082p1246

Lawrie, G. A., Schultz, M., Bailey, C. H., Mamun, A. A., Micallef, A. S., Williams, M., & Wright, A. H. (2016). Development of scaffolded online modules to support self-regulated learning in chemistry concepts. In M. Schultz, T. A. Holme, & S. Schmid (Eds.), *Technology and assessment strategies for improving student learning in chemistry* (pp. 1–21). Washington, DC: American Chemical Society.

Lawrie, G. A., Schultz, M., O'Brien, G., Tasker, R., Bedford, S. B., Dargaville, T. R., … Wright, A. H. (in preparation). *Shared perspectives, strategies and an inter-institutional collaborative community in development of concept items*.

Lawrie, G. A., Schultz, M., & Wright, A. H. (in review). *Insights into students' conceptions as they enter tertiary chemistry studies: A longitudinal study exploring the utility of concept inventories*.

Lawrie, G., Wright, A. H., Schultz, M., Dargaville, T., O'Brien, G., Bedford, S., & Thompson, C. (2013). Using formative feedback to identify and support first-year chemistry students with missing or misconceptions. A practice report. *International Journal of the First Year in Higher Education*, *4*, 111–116. doi:10.5204/intjfyhe.v4i2.179

Lawrie, G. A., Wright, A., Schultz, M., Dargaville, T., Tasker, R., Williams, M., & Thompson, C. (2015). Closing the loop: A model for inter-institutional collaboration through delivering formative assessment in large, first-year stem classes. In G. Weaver, W. D. Burgess, A. L. Childress, & L. Slakey (Eds.), *Transforming institutions: Undergraduate stem education for the 21st century* (pp. 399–410). Purdue, IN: Purdue University Press.

Lewis, E. L., & Linn, M. C. (1994). Heat energy and temperature concepts of adolescents, adults, and experts: Implications for curricular improvements. *Journal of Research in Science Teaching*, *31*, 657–677. doi:10.1002/tea.3660310607

Luxford, C. J., & Bretz, S. L. (2014). Development of the bonding representations inventory to identify student misconceptions about covalent and ionic bonding representations. *Journal of Chemical Education*, *91*, 312–320. doi:10.1021/ed400700q

McInnis, C., James, R., & McNaught, C. (1995). *First year on campus: Diversity in the initial experiences of Australian undergraduates*. Report for centre for the study of higher education, University of Melbourne (Canberra, Australia).

Mulford, D. R., & Robinson, W. R. (2002). An inventory for alternate conceptions among first-semester general chemistry students. *Journal of Chemical Education*, *79*, 739–744. doi:10.1021/ed079p739

Murphy, K., Holme, T., Zenisky, A., Caruthers, H., & Knaus, K. (2012). Building the ACS exams anchoring concept content map for undergraduate chemistry. *Journal of Chemical Education*, *89*, 715–720. doi:10.1021/ed300049w

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*, 199–218. doi:10.1080/03075070600572090

Osborne, R. J., & Cosgrove, M. M. (1983). Children's conceptions of the changes of state of water. *Journal of Research in Science Teaching*, *20*, 825–838. doi:10.1002/tea.3660200905

Othman, J., Treagust, D. F., & Chandrasegaran, A. L. (2008). An investigation into the relationship between students' conceptions of the particulate nature of matter and their understanding of chemical bonding. *International Journal of Science Education*, *30*, 1531–1550. doi:10.1080/09500690701459897

Pavelich, M. J., Jenkins, B., Birk, J., Bauer, R., & Krause, S. (2004). *Development of a chemistry concept inventory for use in chemistry, materials and other engineering courses*. Paper presented at the American Society for Engineering Education annual conference and exposition, Salt Lake City, UT.

Pentecost, T. C., & Langdon, L. S. (2008). *Development of a covalent bonding and molecular structure learning progression and concept inventory for first semester general chemistry*. Paper presented at the 20th biennial conference on Chemical Education, Bloomington, Indiana.

Peterson, R. F., Treagust, D. F., & Garnett, P. (1989). Development and application of a diagnostic instrument to evaluate grade 11 and 12 students' concepts of covalent bonding and structure following a course of instruction. *Journal of Research in Science Teaching*, *26*, 301–314. doi:10.1002/tea.3660260404

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, *66*, 211–227. doi:10.1002/sce.3730660207

Potgieter, M., & Davidowitz, B. (2011). Preparedness for tertiary chemistry: Multiple applications of the chemistry competence test for diagnostic and prediction purposes. *Chemistry Education Research and Practice*, *12*, 193–204. doi:10.1039/C1RP90024B

Potgieter, M., Davidowitz, B., & Blom, B. (2005). *Chemical concepts inventory of first year students at two tertiary institutions in South Africa*. Paper presented at the conference of the South African association of Research in Mathematics, Science and Technology and Education, Namibia.

Potgieter, M., Davidowitz, B., & Venter, E. (2008). Assessment of preparedness of first-year chemistry students: Development and application of an instrument for diagnostic and placement purpose. *African Journal of Research in SMT Education*, Special Edition, *12*, 1–18. doi:10.1080/10288457.2008.10740638

Prince, M., Vigeant, M., & Nottis, K. (2012). Development of the heat and energy concept inventory: Preliminary results on the prevalence and persistence of engineering students' misconceptions. *Journal of Engineering Education*, *101*, 412–438. doi:10.1002/j.2168-9830.2012.tb00056.x

Rasch, G. (1993). *Probabilistic models for some intelligence and attainment tests*. Chicago, IL: MESA Press.

Sadler, P. M. (1998). Psychometric models of student conceptions in science: Reconciling qualitative studies and distractor-driven assessment instruments. *Journal of Research in Science Teaching*, *35*, 265–296. doi:10.1002/(SICI)1098-2736(199803)35:3<265::AID-TEA3>3.0.CO;2-P

Schmidt, H.-J. (1997). Students' misconceptions – looking for a pattern. *Science Education*, *81*, 123–135. doi:10.1002/(SICI)1098-237X(199704)81:2<123::AID-SCE1>3.0.CO;2-H

Schwartz, P., & Barbera, J. (2014). Evaluating the content and response process validity of data from the chemical concepts inventory. *Journal of Chemical Education*, *91*, 630–640. doi:10.1021/ed400716p

diSessa, A. A. (1988). Knowledge in pieces. In G. Forman & P. B. Pufall (Eds.), *Constructivism in the computer age* (pp. 49–70). Hillsdale, NJ: Lawrence Erlbaum Associates.

diSessa, A. A. (2008). A bird's-eye view of the "pieces" vs. "coherence" controversy (from the "pieces" side of the fence). In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 35–60). New York, NY: Routledge.

diSessa, A. A., Gillespie, N. M., & Esterly, J. B. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive Science*, *28*, 843–900. doi:10.1016/j.cogsci.2004.05.003

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*, 107–120. doi:10.1007/S11336-008-9101-0

Singer, S. R., Nielsen, N. R., & Schweingruber, H. A. (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. Washington, DC: National Academies Press.

Solano-Flores, G., & Nelson-Barber, S. (2001). On the cultural validity of science assessments. *Journal of Research in Science Teaching*, *38*, 553–573. doi:10.1002/tea.1018

Solomonidou, C., & Stavridou, H. (2000). From inert object to chemical substance: Students' initial conceptions and conceptual development during an introductory experimental chemistry sequence. *Science Education*, *84*, 382–400. doi:10.1002/(SICI)1098-237X(200005)84:3<382::AID-SCE4>3.0.CO;2-D

Stains, M., Escriu-Sune, M., De Santizo, Molina Alvarez de Santizo, M. L. & Sevian, H. (2011). Assessing secondary and college students' implicit assumptions about the particulate nature of matter: Development and validation of the structure and motion of matter survey. *Journal of Chemical Education*, *88*, 1359–1365. doi:10.1021/ed1002509

Taber, K. S. (2015). Exploring the language(s) of chemistry education. *Chemistry Education Research and Practice*, *16*, 193–197. doi:10.1039/c5rp90003d

Talanquer, V. (2007). Explanations and teleology in chemistry education. *International Journal of Science Education*, *29*, 853–870. doi:10.1080/09500690601087632

Talanquer, V. (2013). How do students reason about chemical substances and reactions?. In G. Tsaparlis, & H. Sevian (Eds.), *Concepts of matter in science education* (pp. 331–346). Dordrecht: Springer Science+Business Media.

Talanquer, V. (2016). Central ideas in chemistry: An alternative perspective. *Journal of Chemical Education*, *93*, 3–8. doi:10.1021/acs.jchemed.5b00434

Tasker, R. (2014). Visualising the molecular world for a deep understanding of chemistry. *Teaching Science*, *60*, 16–27.

Tasker, R., & Dalton, R. (2006). Research into practice: Visualisation of the molecular world using animations. *Chemistry Education Research and Practice*, *7*, 141–159. doi:10.1039/B5RP90020D

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. doi:10.5116/ijme.4dfb.8dfd

Treagust, D. F. (1986). Evaluating students' misconceptions by means of diagnostic multiple choice items. *Research in Science Education*, *16*, 199–207. doi:10.1007/BF02356835

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, *10*, 159–169. doi:10.1080/0950069880100204

Treagust, D. F., Chandrasegaran, A. L., Crowley, J., Yung, B. H. W., Cheong, I. P.-A., & Othman, J. (2010). Evaluating students' understanding of kinetic particle theory concepts relating to the states of matter, changes of state and diffision: A cross-national study. *International Journal of Science and Mathematics Education*, *8*, 141–164. doi:10.1007/s10763-009-9166-y

Treagust, D. F., Chandrasegaran, A. L., Zain, A. N. M., Ong, E. T., Karpudewan, M., & Halim, L. (2011). Evaluation of an intervention instructional program to facilitate understanding of basic particle concepts among students enrolled in several levels of study. *Chemistry Education Research and Practice*, *12*, 251–261. doi:10.1039/C1RP90030G

Tyson, L., Treagust, D. F., & Bucat, R. B. (1999). The complexity of teaching and learning chemical equilibrium. *Journal of Chemical Education*, *76*, 554–558. doi:10.1021/ed076p554

Tytler, R. (2000). A comparison of year 1 and year 6 students' conceptions of evaporation and condensation: Dimensions of conceptual progression. *International Journal of Science Education*, *22*, 447–467. doi:10.1080/095006900289723

Vosniadou, S. (2012). Reframing the classical approach to conceptual change: Preconceptions, misconceptions and synthetic models. In B. J. Fraser, K. G. Tobin, & C. J. McRobbie (Eds.), *Second international handbook of science education* (pp. 119–130). Dordrecht: Springer Science +Business Media.

Vosniadou, S., Vamvakoussi, X., & Skopeliti, I. (2008). The framework theory approach to the problem of conceptual change. In S. Vosniadou, (Ed.), *International handbook of research on conceptual change* (pp. 3–34). New York, NY: Routledge.

Wren, D., & Barbera, J. (2013). Gathering evidence for validity during the design, development, and qualitative evaluation of thermochemistry concept inventory items. *Journal of Chemical Education*, *90*, 1590–1601. doi:10.1021/ed400384g

Wren, D., & Barbera, J. (2014). Psychometric analysis of the thermochemistry concept inventory. *Chemistry Education Research and Practice*, *15*, 380–390. doi:10.1039/c3rp00170a

Yezierski, E. J., & Birk, J. P. (2006). Misconceptions about the particulate nature of matter: Using animations to close the gender gap. *Journal of Chemical Education*, *83*, 954–960. doi:10.1021/ed083p954

Yorke, M. (2001). Formative assessment and its relevance to retention. *Higher Education Research & Development*, *20*, 115–126. doi:10.1080/758483462