Measuring learning gains and examining implications for student success in STEM

Kathleen Scalise\textsuperscript{a}, Michelle Douskey\textsuperscript{b} and Angelica Stacy\textsuperscript{b}

\textsuperscript{a}Department of Educational Methodology, Policy and Leadership, University of Oregon, Eugene, OR, USA; \textsuperscript{b}Department of Chemistry, University of California, Berkeley, CA, USA

\textbf{ABSTRACT}

This article focuses on learning gains in the area of higher education STEM (science, technology, engineering, and mathematics) through new approaches to the formative assessment feedback cycle. Learning gains were examined for stratified randomized comparison and two treatment groups of 943 students organized into 41 laboratory sections in first-year chemistry. Treatment groups received either (i) support for conceptual understanding through a larger percentage of conceptual questions in online homework or (ii) the same mix of conceptual questions with differentiated answer feedback. The comparison group received the usual online homework and feedback for the course. Results showed no statistically significant difference among groups at pretest, yet both treatment groups showed statistically significant gains in learning outcomes over the original comparison group. Additional findings included (i) students in the treatment condition showed a 43\% reduction in STEM ‘at risk’ identification by mid-course and (ii) a gender effect was identified.

\textbf{Introduction}

This case study reports on one approach to assess student learning outcomes at the post-secondary level, using learning trajectories in STEM courses (science, technology, engineering, mathematics). Learning trajectories, sometimes called learning progressions, are ways to describe pathways of student learning over a course of instruction, where learning goals have been described and will be measured in a cognitive framework (Corcoran, Mosher, & Rogat, 2009; Wilson, 2009).

Research into a variety of measures of learning gain for higher education has been gaining traction (van Barneveld, Arnold, & Campbell, 2012; Barr & Matsui, 2007; Claesgens, Scalise, Wilson, & Stacy, 2008; U.S. Department of Education, 2006; Wilson & Scalise, 2006). The effort here was launched because post-secondary success in mathematics and science courses (Johnson, Arumi, & Ott, 2006) is an important gatekeeper to many attractive careers, is important for life skills, and can be key to a successful economy.
In the U.S., student preparedness in their undergraduate STEM course experiences can determine whether such career options are open to them (Scalise, Claesgens, Wilson, & Stacy, 2006). For instance, a longitudinal study of incoming first-year students at two U.S. universities, UC Berkeley and Stanford University, indicated a drop in interest over time for medical careers following participation in STEM post-secondary courses (Barr & Matsui, 2007). In this study, the principal reason cited by students for their loss of interest in the medical field was their experience in one or more chemistry courses.

Understanding what students know and how they come to know it can be key to supporting learning and success in such courses (Black, Harrison, Lee, Marshall, & Wiliam, 2002; National Research Council, 2001). Scholars argue that students and instructors need effective formative assessment, or assessment for learning, rather than only of learning, to inform numerous objectives, including student feedback, teacher feed-forward, the development of cognitive engagement (Black et al., 2002; Black, Harrison, Lee, Marshall, & Wiliam, 2003; Black & Wiliam, 1998).

The three research questions for the study are: (i) whether it is possible to assess student knowledge in a post-secondary STEM course relative to a conceptual framework, (ii) if learning trajectories, or pathways of student learning framed as learning progressions, can be successfully employed with latent variable methods, and (iii) how the evidence might be useful for instructional decision-making, for instance, to better support successful learning outcomes in post-secondary education.

**Research literature latent variables, constructs and frameworks**

Describing what we want students to learn and what successful learning patterns should look like for particular learning objectives is important for assessments (Wilson, 2005). One approach is to capture not only what mastery should look like but also some patterns of understanding students exhibit as they move toward mastery. To express what we want to measure, we can describe a latent variable, or set of latent variables, as a framework structure on which to assess students, such as based on an approach used here from the University of California, Berkeley BEAR Assessment System (Wilson, 2005).

One of the challenges in science education, as in many other areas of instruction, is that learning objectives and goals often describe the body of knowledge to be mastered in discrete portions or separate isolated standards (Darling-Hammond, 2004). The knitting together, or relationships among the knowledge and concepts, are very important to the development of each student’s understanding, as has been addressed by many educational standards documents (American Association for the Advancement of Science, 1993; National Research Council, 1996).

For instance, chemistry is a set of powerful models of the natural world for which the discrete knowledge (Smith, Wiser, Anderson, & Krajcik, 2006) are intended to build understanding. The aim is to build working knowledge of aspects of these powerful scientific models, as students show evidence of coordinating this knowledge into a functional whole. Without this view, the focus of instruction can become a fragmented acquisition of facts and algorithms rather than development of an integrated knowledge structure (Bodner & Klobuchar, 2001; Driver, Asoko, Leach, Mortimer, & Scott, 1994; Hesse & Anderson, 1992). A conceptual framework can help explicitly describe these important relationships, even
though the proposed latent variables may not be a unique set or completely independent from each other.

‘Perspectives of Chemists’ conceptual framework: an example

In this example, we use a conceptual framework called the Perspectives of Chemists (Claesgens, Scalise, Draney, Wilson, & Stacy, 2002; Claesgens, Scalise, Wilson, & Stacy, 2009; Scalise et al., 2004). The framework was originally developed in a curriculum project funded by the U.S. National Science Foundation. It describes learning within three scientific models: matter, change, and energy (Claesgens et al., 2002). A high-level overview of the three variables in the Perspectives framework is shown in Figure 1.

The Perspectives are built on the idea that progressions of student learning in introductory chemistry can be largely grouped into these three core conceptions, as scientific models about matter, change, and energy. The purpose in framing the big ideas of chemistry is to organize the overarching ideas of the discipline while simultaneously constructing an instrument for measuring the values of these variables for each student.

In addition to this organization of topics, the Perspectives further suggest that learning in each of these three areas or strands can be modeled as a progressive continuum from a naive to a more complete and normative understanding of explanatory models of chemistry, and that these continuums represent a set of necessary and distinct areas of understanding enabling higher order thinking in chemistry. Advancement through the levels is designed to be cumulative. The emphasis is on understanding, thinking, reasoning, and using practices with chemistry that relate basic concepts (ideas, facts, and models) to analytical reasoning.

Within each of the Perspectives, a five-level scale is used to describe student understanding. The levels within the proposed variables are constructed such that students are required to give more complex and sophisticated responses to increase their score from describing their initial ideas in Level 1 (Notions), to relating the language of chemists to their view of the world in Level 2 (Recognition), to formulating connections between several ideas in Level 3 (Formulation), to fully developing models in Level 4 (Connection), to asking and researching new questions in Level 5 (Extension).

Empirical data are then collected and item response modeling used to analyze and interpret the data (Wilson, 2005). Item response models are one form of latent variable analysis, and they are used here to calibrate the framework. There is an accumulating body of evidence to support use of item response model measurement approaches in the classroom.

| Perspectives of Chemists |
|-------------------------|
| Matter                  | Matter is composed of atoms arranged in various ways. |
| Change                  | Change is associated with the rearrangement of atoms. |
| Energy                  | Energy is associated with changes that occur. |

Figure 1. Three of the Perspectives of Chemists variables.
(De Boeck & Wilson, 2004; Embretson, 2000; Hambleton, Swaminathan, & Rogers, 1991; Kennedy, Wilson, Draney, Tutunciyan, & Vorp, 2006). Also, computer technology has been developed to expedite classroom use for the creation and validation of assessments based on cognitive theory (Kennedy et al., 2006).

**Methods**

**Study design**

A three-group randomized controlled trial (RCT) design was selected for the study, with pretest, mid-course, and post-course assessments. Treatment and comparison conditions for the three groups consisted of (i) using a standard publisher’s homework package used previously for the course with question feedback (COMPARISON condition), (ii) substituting some more conceptual questions into homework sets for students (CONCEPTUAL condition) and (iii) adding to this differentiating instruction by providing feedback on Perspectives level (DIFFERENTIATED condition).

The intervention in the CONCEPTUAL focused on replacing some of the standard questions in the weekly homework sets. The questions were provided to the course as part of the adoption of an electronic textbook. Questions were evaluated by instructors and for the CONCEPTUAL condition, some of the questions were replaced each week to focus on problem solving that involved a stronger emphasis on conceptual thinking. The intervention in the DIFFERENTIATED employed the same modified homework sets as in the CONCEPTUAL condition, but with additional feedback provided to students to describe their progress in the learning progressions. The comparison group received the usual online homework activities and feedback for the course, which consisted of the same number and length of problem sets but with less conceptual focus and standard worked solutions identical for all students.

Randomization was conducted in the first week of instruction. Students received a pretest and were stratified by lab section according to pretest results. Sections were then matched by stratification and first randomly assigned half to treatment condition and half to the comparison condition. Finally, sections in the treatment condition were then randomly assigned to one or the other of the two treatment conditions.

As part of the human subjects protocol, after the first seven weeks of instruction (at the end of Time 1), students in the comparison condition were moved to a treatment condition, and students in the treatment condition were moved to the comparison condition. This is called exchanged. After the second seven weeks (Time 2), all students received a final set of assessments. Therefore over the 14 weeks of instruction, all students participated in both the comparison and one of the treatment conditions, over Time 1 and Time 2. Note that this study design does have some limitations, especially involving the single case site and limitations of the sample available, see limitations listed in the Results and Discussion section later in the paper.

**Treatments**

All students across conditions received the same total number of homework questions, and on the same topics, throughout the course. About 15% of the homework questions in
each of approximately four-weekly homework assignments for each of Time 1 and Time 2 were adjusted to have a more conceptual focus for the CONCEPTUAL condition. Note that students in all conditions at both time periods also received many conceptual and other questions throughout the course, in preclass and prelab assignments, in-class i-clicker sessions, lecture prompts, and laboratory sessions.

All students regardless of condition received immediate online feedback on their homework answers with worked solutions, but those in the DIFFERENTIATED condition also received four short feedback emails over the course of the seven weeks when in the treatment condition, describing some aspects of strategy improvement directly related to the conceptual framework, *Perspectives of Chemists*. Note that no treatment consisted of DIFFERENTIATED condition alone without the CONCEPTUAL condition. This is because in this case, the differentiated feedback was designed in response to the conceptual framework and therefore applied to questions in the CONCEPTUAL condition. Therefore, it was not possible to supply the DIFFERENTIATED feedback without first supplying the CONCEPTUAL condition.

**Linking assessments**

In order to link the assessments over time in the course, statistical techniques were used. Linking was completed using overlap items from the item banks during the item calibration. Common items were administered and used at pretest, at the end of Time 1, and at the end of Time 2. Link items were calibrated together for the linking and represented approximately 20% of the total items administered.

**Analysis**

For analysis, student proficiency estimates were made using the item response model to generate scores for each students at each of the time points, calibrated to the *Perspectives* framework. Then, one-way ANOVA was conducted first on pretest results, to compare the baseline performance at start of study across groups. Next, a three-group one-way ANOVA was conducted to compare the mid-course results at the end of Time 1, followed by further analysis at the end of Time 2. Results are reported in the next section.

**Results and discussion**

**Sample**

The sample group consisted of 973 students in their first post-secondary course in general chemistry at UC Berkeley. Students were enrolled in 41 lab sections of the same course. Internal review board (IRB) approvals were obtained from the campus, which is the standard U.S. protocol for human subjects approval when research involves students in courses. Students of age 18 or older could consent to have data included in analysis.¹ Students following placement into group did not know their placement status, but all students participated for an equal time in treatment and comparison groups.
**Instrumentation and model fit**

Regarding instrumentation and model fit for the calibration of the instruments into a linked item bank, fit statistics were generally acceptable, with 96% of items within typical values acceptable for fit to the model employed. Overall, test reliability results for the final time point was reasonably good at MLE Person Separation Reliability of .82 and EAP/PV reliability of .85.²

**Pretest results**

Regarding the pretest results across groups, there was no significant difference at pretest among the groups (p = .35). The original groups therefore are considered equivalent in performance at the start of the study.

**Time 1 results**

For the three-group one-way ANOVA conducted to compare the raw score of the mid-term exam results at the end of Time 1, results rejected the null and supported the hypothesis that the different treatment conditions were associated with differential learning outcomes, $F(3,828) = 3.85$, $p < .01$. Post-hoc-independent sample $t$ tests revealed a significant difference between COMPARISON and CONCEPTUAL conditions ($p < .01$) and between COMPARISON and DIFFERENTIATED conditions ($p < .01$). The mean mid-term exam score on average was 4 points higher for each of the CONCEPTUAL and DIFFERENTIATED conditions.

**Time 2 results**

At Time 2, there was no hypothesis in advance of which group or groups if any were expected to be higher performing at the end of Time 2, given the possibility of both ordering effects and efficacy of the interventions themselves. At the final proficiency estimate (Week 15), two-tailed $t$ tests were significant, $p = .02$. Students in the original DIFFERENTIATED and CONCEPTUAL conditions at Time 1 ($M = .22$, SD = .58) continued to significantly outperform the students who at Time 2 received the CONCEPTUAL treatment ($p = .02$, $M = .10$, SD = .58) or DIFFERENTIATED treatment ($p = .02$, $M = .11$, SD = .58). There was no significant difference between the two new treatment groups at the end of T2 ($p = .22$). However, note that a trend existed between the two treatment groups at Time 1, with the DIFFERENTIATED treatment at Time 1 significantly higher than CONCEPTUAL condition, although the difference did not persist into Time 2.

Results of the comparisons are summarized in Figure 2, which shows student proficiency outcomes at three time points by treatment condition (CP is COMPARISON, T-CO is the CONCEPTUAL, and T-DI is the DIFFERENTIATED treatment group). Raw scores on the final examination reflected a persisting Cohen's $d$ effect size of about .2 standard deviations between those in the original treatment conditions compared to those in the original comparison conditions.
Perspectives framework and learning trajectory interpretations

To interpret results as trajectories on the Perspectives of Chemists framework, here we focus on the critical shift from ‘Notions’ thinking, in which students respond across numerous questions with answers that are generally incorrect or out of scope, to the next level, ‘Recognition,’ where students are using chemistry models accurately enough that they are gaining some successful problem-solving power from their new understandings.

We see this move from Notions to Recognition as a critical shift in conceptual understanding for introductory general chemistry. The Recognition level provides students with a conceptual foundation on which correct understanding of models can begin to build. Students who remain in Notions are not yet knitting together the ideas and relationships sufficiently to begin to use chemistry models to add any power as conceptual support for tasks and activities.

Figure 3 shows the distribution over the Perspectives framework for all students shown by time point (pre, mid, post). ‘Notions’ is scored at the 1 level and is represented here by the 1−, 1, and 1+ estimates, with minus and plus representing heading into and out of the level. ‘Recognition’ is scored at the 2 level and is represented here by the 2−, 2, and 2+ estimates. As can be seen in Figure 3, few students reached above the Recognition level by the end of this introductory course, which is not unexpected as knowledge would extend in subsequent courses.

To interpret Figure 3, first at pretest, essentially all students in the data-set responded on average with Notions-type answers. Some students reasoned with logic and real-world knowledge at the 1 level, and others added attempts at using chemistry concepts to reach
the 1+ level, but regardless nearly all answers were essentially incorrect or out of scope, on the pretest prior to instruction, for this data-set, which again is not unexpected as students have not yet had the intended instruction.

By mid-point in the course, after approximately seven weeks of instruction, many of the general chemistry students showed a substantial shift toward 'Recognition' in their responses. This can be seen in Figure 3 in the 'mid' bars, where students have mostly shifted forward from the 1 level, to 1+, 2−, and 2.

By the post-course assessment, the shift to the right on Figure 3 has continued, with more students moving into 'Recognition' and leaving 'Notions' thinking behind. It is key to note that even by the end of course, however, many students remained responding across numerous tasks mostly with 'Notions' thinking. These students would be expected to be at risk in subsequent chemistry courses, which could be a question for future investigation.

Next, Figure 4 shows the distribution at the post-assessment time point by the original treatment/comparison groups. The statistically significant difference reported previously is now seen here in Perspectives levels. Here it can be seen that substantially more students from the original COMPARISON group remained answering with 'Notions' thinking by the end of the course, as compared to the original treatment groups of CONCEPTUAL or DIFFERENTIATED.

Next, we consider Perspectives results by gender, with gender self-reported by students. In this data-set, and in others we have looked at for this course, there is a statistically significant gender effect in the course outcomes, with males outperforming females. For this data-set, there was a slight difference in Perspectives level at pretest, with males slightly higher. However, males in this data-set outperformed females on average by 20 points on the post-course assessment, with male scores approximately considerably higher at 15% above female ($p < .01$). Figure 5 shows more women remaining in 'Notions' thinking by
end of course, while a higher percentage of males are moving into Recognition, indicating firmer conceptual grasp for these students.

While the data-set here does not include research questions to investigate the gender-related results found, the instructors who engaged in the study, both female, anecdotally reported two factors that might be investigated in future studies. First, the instructors found male students in the course tended to structure their question-asking activities in class and lab as verification questions. In verification, students assemble their conception

**Figure 4.** Distribution of students on final proficiency by original treatment.

**Figure 5.** Distribution of general chemistry students on final proficiency by gender.
of a model and seek verification in their STEM work, an active strategy of questioning. Female students seemed to be asking questions in ways that were more passive regarding conceptual model building. Questions in a more passive style require others to construct the model and inform on it, without the questioner having previously engaged in the prior model building efforts that a verification question requires.

Furthermore, the instructors reported male students who felt challenged by the material after completing the readings and assignments seemed to seek out more problem-solving activities in outside resources, such as through other online chemistry sites not associated with the course. Their female students seemed to the instructors more likely to report returning to re-read the original sources, which may not have extended knowledge to the same degree as consulting a new source. Neither the verification question strategy nor the outside resource strategies were a class topic.

While such strategies are not the focus of this study, interventions that focused on supports in these directions could provide utility to gauge effects; measuring learning outcomes on a conceptual framework such as the Perspectives could help shed light and better provide answers to gaps seen such as these, if the intervention is explored.

Finally, Figure 6 shows Perspectives results for 100 students in the course identified as most at-risk of credit denial by the mid-point of the course, based on their point accumulation. The vast majority of the at-risk students remained in ‘Notions’ by the end of the course, and their work reflected an absence of much conceptual understanding. For students in the original CONCEPTUAL treatment group, however, the rate of at-risk identification by mid-point was significantly lower, at 8.5% as compared to the original COMPARISON condition, at 14.8%. Thus the treatment is associated here with a substantial 43% reduction of risk for outcomes, such as credit denial in the course, and inability to proceed in a selected pathway as STEM students.

Figure 6. Distribution of general chemistry students on final proficiency by at-risk.
Limitations of the study

Limitations of this study are numerous. First of all, as a single case study design, one study site is not representative. Secondly, from a measurement science perspective, while as briefly mentioned models fit reasonably well and indicated acceptable reliability for the instruments, often a sequence of models are explored for best fit. This could be done in future work.

Applying multilevel models to the data-set would also be most appropriate, with variance considered at the lab group level as well as the individual level. However, a larger multi-case study design would be needed. The sample size limits the approach and is a factor to treat the results with caution. Also, more treatment groups might have been desirable, to include a group with differentiated feedback but no higher percentage of conceptual questions. Again, this was not possible due to sample size, and undermines much interpretation of the differentiated condition as compared to the conceptual, although both made similar gains.

Finally to mention here, continuing with the original treatment and comparison groups from the RCT throughout the entire course rather than exchanging at mid-point could have been desirable, to extend the length of the treatment. However, this was not possible due to ethical concerns for human subjects, given the nature of the course-embedded formative assessments.

Conclusion

Overall, the RCT findings showed that the treatment groups receiving in Time 1 (i) conceptual support or (ii) conceptual support with differentiated instruction showed a statistically significant gain in learning outcomes over the comparison group. These original treatment group on average continued to outperform the original comparison group even when the groups were exchanged at mid-point. This suggests that conceptual support early on may have sustained impacts on understanding over time.

Additional findings are that students in the conceptual condition showed a 43% reduction in those identified as at risk in their chemistry studies by course mid-point, as compared to the comparison group. Also, a persistent gender effect was detected throughout the time points. A slightly higher Perspectives level by male students in the sample group at pretest grew more substantial over time, regardless of treatment group, resulting in an approximate 15% difference on final exam scores and a substantially different spread in Perspectives levels for men and women by the end of the course.

We conclude regarding the three research questions that based at least on this case study, it can be possible to assess student knowledge in a post-secondary STEM course relative to a conceptual framework, such as shown here drawing on the BEAR assessment system approach with latent variable models. Secondly, a type of learning trajectory was here successfully identified with the methods relative to the framework. Models fit reasonably well, instruments showed reasonable reliability, and results were interpretable. Finally, regarding how the evidence might be useful for instructional decision-making, it was seen that the interpretation of change over the trajectory of the framework revealed some useful information regarding student learning patterns, including how the treatment conditions were associated with improved learning gains overall, and the generation of results for at-risk and gender gap comparisons. We conclude that focusing on learning gains in the area of higher education STEM (science, technology, engineering, and mathematics) through such
approaches as shown here may provide some utility, and may offer some new approaches to considering both student learning gains and the efficacy of new interventions.

Notes

1. 1263 cases total included a proficiency estimate possible at pretest, mid-term and/or post-test. Of these 77 did not respond yes or no on consent, so were removed. An additional 10 declined consent for their data to be included in the study, and were removed. Four provided consent but declined to state whether 18 years or older so were removed. An additional 199 students were not age 18 or older at the time of original age data collection and so were removed. The remainder of 973 students provided data for use in this analysis.

2. Due to systematically missing data by design, Cronbach's alpha was not generated but rather the reliability indices reported above were used.

Acknowledgments

The authors thank Jennifer Claesgens, Paul Daubenmire, Rebecca Krystyniak, Tatiana Lim, Sheryl Mebane, Sandhya Rao, Michelle Sinapuelas, and Nathaniel Brown for their assistance with instrument and framework development, data collection, scoring, and assistance in discussions of student learning patterns.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the U.S. National Science Foundation [grant number DUE: 0125651].

ORCID

Michelle Douskey http://orcid.org/0000-0002-3942-5925

References

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York, NY: Oxford University Press.

Barr, D.A., & Matsui, J. (2007). *The “turning point” for minority pre-meds: The effect of early undergraduate experience in the sciences on aspirations to enter medical school of minority students at UC Berkeley and Stanford University*. Research & Occasional Paper Series: CSHE.20.08. Berkeley: Center for Studies in Higher Education, University of California.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2002). *Working inside the black box: Assessment for learning in the classroom*. London: King's College.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Buckingham: Open University Press.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–148.

Bodner, G., & Klobuchar, M. (2001). The many forms of constructivism. *Journal of Chemical Education*, 78(8), 1107–1134.

Claesgens, J., Scalise, K., Draney, K., Wilson, M., & Stacy, A. (2002). *Perspectives of Chemists: A framework to promote conceptual understanding in chemistry*. Paper presented at the Validity and
Value in Educational Research American Educational Research Association Annual Meeting, New Orleans, LA.
Claesgens, J., Scalise, K., Wilson, M., & Stacy, A. (2008). Assessing student understanding in and between courses for higher education: An example of an approach. Assessment Update, 20(5), 6–8.
Claesgens, J., Scalise, K., Wilson, M., & Stacy, A. (2009). Mapping student understanding in chemistry: The Perspectives of Chemists. Science Education, 93(1), 56–85.
Corcoran, T., Mosher, F.A., & Rogat, A. (2009). Learning progressions in science: An evidence-based approach to reform. New York, NY: Center on Continuous Instructional Improvement, Teachers College – Columbia University.
Darling-Hammond, L. (2004). Standards, accountability, and school reform. Teachers College Record, 106(6), 1047–1085.
De Boeck, P., & Wilson, M. (2004). Explanatory item response models: A generalized linear and nonlinear approach. New York, NY: Springer.
Driver, R., Asoko, H., Leach, J., Mortimer, E., & Scott, P. (1994). Constructing scientific knowledge in the classroom. Educational Researcher, 23(7), 5–12.
Embretson, S. (2000). Item response theory for psychologists. Mahway, NJ: Lawrence Erlbaum Associates.
Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). Fundamentals of item response theory. Newbury Park, CA: Sage.
Hesse, J., & Anderson, C.W. (1992). Students’ conceptions of chemical change. Journal of Research in Science Teaching, 29(3), 277–299.
Johnson, J., Arumi, A.M., & Ott, A. (2006). Reality check 2006, issue no. 3: Is support for standards and testing fading? Public Agenda, Retrieved from http://www.publicagenda.org/research/pdfs/rc0603.pdf
Kennedy, C., Wilson, M.R., Draney, K., Tutunciyan, S., & Vorp, R. (2006). GradeMap v4.2 user guide home page. Berkeley, CA: Berkeley Evaluation and Assessment Research Center, University of California.
National Research Council. (1996). National science education standards. Washington, DC: National Academies Press.
National Research Council. (2001). Knowing what students know: The science and design of educational assessment. Washington, DC: National Academies Press.
Scalise, K., Claesgens, J., Krystyniak, R., Mebane, S., Wilson, M., & Stacy, A. (2004). Perspectives of Chemists: Tracking conceptual understanding of student learning in chemistry at the secondary and university levels. Paper presented at the Enhancing the Visibility and Credibility of Educational Research, American Educational Research Association Annual Meeting, San Diego, CA.
Scalise, K., Claesgens, J., Wilson, M., & Stacy, A. (2006). Contrasting the expectations for student understanding of chemistry with levels achieved: A brief case-study of student nurses. Chemistry Education Research and Practice, The Royal Society of Chemistry, 7(3), 170–184.
Smith, C.L., Wiser, M., Anderson, C.W., & Krajcik, J. (2006). Implications of research on children’s learning for standards and assessment: A proposed learning progression for matter and the atomic molecular theory. Measurement: Interdisciplinary Research and Perspectives, 4(1 & 2), 1–98.
U.S. Department of Education. (2006). A test of leadership, charting the future of U.S. higher education. A report of the commission appointed by Secretary of Education Margaret Spellings. Retrieved from http://www.ed.gov/about/bdscomm/list/hiedfuture/index.html
van Barneveld, A., Arnold, K.E., & Campbell, J.P. (2012). Analytics in higher education: Establishing a common language. EDUCAUSE Learning Initiative. Retrieved from https://qa.itap.purdue.edu/learning/docs/research/EL13026.pdf
Wilson, M. (2005). Constructing measures: An item response modeling approach. Mahwah, NJ: Lawrence Erlbaum Assoc.
Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. Journal of Research in Science Teaching, 46, 716–730.
Wilson, M., & Scalise, K. (2006). Assessment to improve learning in higher education: The BEAR Assessment System. Higher Education, 52, 635–663.