

*John B. Bond, David W. Denton & Arthur K. Ellis (USA)*

## **Impact of Reflective Assessment on Student Learning: Best-Evidence Synthesis from Ten Quantitative Studies**

**Summary:** *Formative assessment involves feedback to teachers for informing instruction and also feedback to students for directing their own learning. Early research on formative assessment showed independence from any particular theoretical foundation. Self-regulated learning theory provides a helpful construct for organizing formative assessment through familiar classroom practices, including provision of feedback, strategy use, and metacognition. One way to integrate reflective activities is with reflective assessment, which emphasizes gathering feedback through questioning, writing, and discussing. Ten studies were analyzed using best-evidence methodology to show the effects of reflective assessment on student performance of posttest and retention tests. Weighted mean effect sizes ranged from .28 to .37. Results suggest additional investigations into the use of reflection for improving student learning and other outcomes.*

**Keywords:** *best-evidence synthesis, effect size, feedback, formative assessment, metacognition, reflective assessment, self-regulated learning, strategy use*

**Резюме** (Джон Б. Бонд, Давид В. Дентон & Артур К. Эллис: Воздействие рефлексивной оценки на обучение учеников: синтез убедительного доказательства из десяти количественных исследований): *Формирующая оценка включает обратное сообщение учителям для организации урока и ученикам для организации их собственного обучения. Прежние исследования по вопросам формирующей оценки показывают независимость от любых специальных теоретических оснований. Теория самоуправляемого обучения предлагает полезную концепцию для организации формирующей оценки с помощью известных практик обучения, включая обратную связь, использование стратегии и метакогницию. Способ интегрирования рефлектирующих видов деятельности – это рефлектирующая оценка, подчеркивающая сочетание обратной связи, устного и письменного опроса и дискуссии. С использованием методики убедительного доказательства были проанализированы десять исследований, чтобы показать воздействие рефлектирующей оценки на достижения учеников при проведении заключительного теста после завершения курса и теста на проверку способности запоминания. Средние показатели величины эффекта составили от 0,28 до 0,37. Результаты говорят о необходимости проведения дальнейших исследований об использовании рефлексии для улучшения успехов в учебе и прочих достижениях.*

**Ключевые слова:** *синтез убедительного доказательства, величина эффекта, обратная связь, формирующая оценка, метакогниция, рефлектирующая оценка, самоуправляемое обучение, использование стратегии*

**Zusammenfassung** (John B. Bond, David W. Denton & Arthur K. Ellis: Auswirkungen der reflexiven Beurteilung auf das Lernen der Schüler: Eine Best-Evidence Synthese aus zehn quantitativen Studien): *Die formative Beurteilung beinhaltet eine Rückmeldung an die Lehrer für die Unterrichtsgestaltung sowie für die Schüler hinsichtlich der Ausrichtung ihres eigenen Lernens. Frühere Forschungen zur formativen Beurteilung zeigten die Unabhängigkeit von jeglichen besonderen theoretischen Grundlagen. Eine Theorie des selbstregulierten Lernens bietet ein hilfreiches Konstrukt für die Organisation formativer Beurteilungen durch vertraute Unterrichtspraktiken, einschließlich des Feedbacks, des Strategie-Einsatzes und der Metakognition. Ein Weg, reflektierende Aktivitäten zu integrieren, ist die reflektierende Bewertung, die die Zusammenführung von Feedback, mündlicher und schriftlicher Befragung und Diskussion hervorhebt. Zehn Studien wurden unter Verwendung der Best-Evidence-Methodik analysiert, um die Auswirkungen der reflektierenden Beurteilung auf die Schülerleistungen bei Posttest und Retentionstests zu zeigen. Die gewich-*

teten mittleren Effektstärken reichten von 0,28 bis 0,37. Die Ergebnisse sprechen für weitere Untersuchungen über die Verwendung von Reflexion zur Verbesserung der Lernerfolge der Schüler und weitere Erkenntnisse.

**Schlüsselwörter:** Best-Evidence-Synthese, Effektstärke, Feedback, formative Beurteilung, Metakognition, reflektierende Bewertung, selbstgesteuertes Lernen, Strategie-Einsatz

## Introduction

The literature base supporting formative assessment is substantial and increasing. However, early research by Black and Wiliam (1998) distinguishes itself from the broad array of theoretical and empirical articles currently available to practitioners and academics. One reason for this is the way Black and Wiliam worked to explain underlying tenets of effective assessment for improving achievement, which were sensible and straightforward, and which aligned with the practical experience of most educators. One of the conclusions reached by Black and Wiliam is that formative assessment involves feedback to teachers for informing instruction. Another is that students themselves gather feedback for directing their own learning and for correcting errors. And yet another is that formative assessment depends on actively engaging students, which is widely accepted as a principle of effective instruction appropriate to all levels and disciplines. Likewise, though not explicitly stated by Black and Wiliam but rather inferred, is the idea that assessment need not always be linked to evaluation. Bloom, Hastings, and Madaus (1971) detached assessment from evaluation a decade earlier, reinforcing the notion that feedback from students be used for day-to-day adjustments made by teachers and students at the classroom level.

An important characteristic of the early research compiled by Black and Wiliam (2009) was its independence from any particular theoretical foundation. As they indicated, formative assessment “did not start from any pre-defined theoretical base but instead drew together a wide range of research findings relevant to the notion of formative assessment” (Black & Wiliam, 2009, p. 5). While excluding theory from research has disadvantages (Knowles, 1990), in the case of formative assessment, it proved helpful. The absence of unifying theory encouraged researchers to explore a variety of instructional practices generally thought to be formative, along with distinctive theories for justifying their use. As a result, formative assessment is wide-ranging and has been applied to various categories of teacher and student activity, including teacher-made-observations, discussion, questioning, graphic organizers, and student self-assessment, among others. Later, Black and Wiliam (2009) provided their own unifying theory of formative assessment, but before they had, researchers were already associating it with principles of curriculum design, teacher to student interdependence, peer to peer interdependence, classroom discourse, mastery learning, and self-regulated learning, among other concepts, constructs, and models.

Although there are several theories that researchers have linked to formative assessment, self-regulated learning significantly broadens the possibilities of improving classroom practice through teacher and student efforts. Consideration of self-regulated learning components - including motivation, metacognition, and behavior - contribute to a nuanced understanding of formative assessment, and also one that is complex and integrative of multiple fields. Part of the complexity has to do with assumptions underlying self-regulated learning. Self-regulation assumes learners are agents who construct knowledge and that all learners self-regulate, but with varying degrees of precision and efficiency (Winne, 2005). One of the intriguing implications of these assumptions is that students can improve their capacity for self-regulation through scaffolding activities, or strategies, such as provision of feedback.

Indeed, feedback is one of the defining characteristics of both formative assessment activities and

self-regulated learning theory. For example, Black and Wiliam (2009) suggest “feedback is a critical feature in determining the quality of learning... and is therefore a central feature of pedagogy” (p. 6). Moreover, according to Black and Wiliam, feedback is not just gathered by teachers for modifying instruction, it is also gathered by students for selecting a strategy or changing a behavior. With respect to self-regulation, Winne (2005) indicates learners need feedback to understand whether their efforts are producing desired results (p. 562). Likewise, Zimmerman (1989) emphasizes the importance of feedback for regulating motivation and behavior, along with the use of specific learning strategies that enable students to monitor whether their efforts are producing improved outcomes.

Strategy use is another definitive characteristic for each field. A strategy is any procedure applied for accomplishing an academic task (Pressley & Harris, 1990). Major elements of teaching strategies to students include (a) demonstrating the strategy in the context of a meaningful academic task, (b) introducing strategies one at a time, (c) providing feedback and opportunities for practice, and (d) assisting students that struggle with the strategy on an individual basis (Pressley & Harris, 1990). Similar to feedback, researchers have included strategy use as a significant component of formative assessment and self-regulated learning. For example, Black and Wiliam (2009) suggest, “feedback on understanding of the task may have to be linked with feedback on the learner’s understanding of the criteria used in his/her own self-regulation, or on the choice of strategy made in the light of that understanding” (p. 24). Likewise, Zimmerman (1989) has suggested a students’ self-regulative knowledge is dependent on the application of a strategy and feedback from its use (p. 332).

A third definitive characteristic of formative assessment and self-regulated learning is metacognition. Flavell (1976) defines metacognition as heightened awareness of one’s thought processes, or “knowledge concerning one’s own metacognitive processes or anything related to them” (p. 232). Zimmerman (2002) situates metacognition within self-regulated learning, and suggests metacognition involves several cognitive skills including (a) setting goals, (b) adopting strategies, (c) evaluating the efficacy of one’s methods, and (d) adapting future methods. Similarly, Dignath and Büttner (2008) add that metacognition includes planning the completion of a task, monitoring one’s comprehension through self-testing, and evaluating one’s learning products in comparison to a goal. Dignath and Büttner also emphasize the importance of teachers communicating to students how, when, and where to apply various metacognitive strategies while also illustrating the benefits of their use.

In summary, characteristics of formative assessment make it amenable to a variety of learning theories and instructional practices. Self-regulated learning theory provides a helpful construct for organizing formative assessment through familiar classroom practices, including provision of feedback, strategy use, and metacognition. Black and Wiliam (1998) themselves justified these connections through their early definition of formative assessment, which they reported as any activity undertaken by “teachers – and by their students in assessing themselves...[to] provide information to be used as feedback” (p. 140).

## **Reflective Assessment**

While provision of feedback, strategy use, and metacognition tell how to unify formative assessment as a possible expression of self-regulated learning theory, these fields include their own questions of how they are implemented at the classroom level. One way to improve coherence is by focusing on a finite set of learning activities, such as those identified as reflective assessment. Reflective assessment emphasizes gathering feedback through observing, questioning, writing, illustrating, and discussing. Information gathered through reflection is intended for use by both teachers and students. A few reflective assessment strategies follow for illustration.

*I learned* statements are comments spoken or written by students summarizing whatever they learned from the lesson (Ellis, 2001, 2010). There are various ways to implement *I learned*, such as having students share their thinking with nearby peers, or writing an Exit Slip. Questions for eliciting *I learned* statements include

- What did you learn?
- What part of the lesson did you find most interesting?
- What is the value of what you learned?
- What do you think you will remember from today's lesson?

A strategy similar to *I learned* is *key idea identification* (Ellis, 2001, 2010), which depends on broader unit goal statements, sometimes referred to as the unit focus, central focus, guiding question, essential question, big idea, or concept. Questions for eliciting key idea identification from students include

- How does yesterday's lesson relate to today's lesson?
- How do you summarize what you have learned from these last few days?
- What is the key idea that explains our activities over the last few weeks?

Another strategy is *clear and unclear windows*, which uses comparisons, rather than lesson or unit goals, as its subject matter (Ellis, 2010). According to Marzano, Pickering, and Pollock (2001) making comparisons is an effective form of instruction, and also flexible since comparisons are readily shown visually, such as Venn diagrams, tables, and graphs. T-charts are yet another visual method for comparing two or more characteristics of things. Marking one side clear and another side unclear turns the chart into clear and unclear windows. Students use the chart for identifying parts of the lesson that make sense and those that are confusing.

Similar to clear and unclear windows with respect to its visual characteristics, *learning illustrated* (Ellis, 2010) focuses on reflection through images, pictures, diagrams, and other representations that are readily understood by students, especially since most brain activity is occupied with processing visual information (Medina, 2008). Some prompts for eliciting illustrations include

- What picture can you draw to show your learning?
- Summarize your learning by illustrating a graphic organizer.
- How can you represent this information as a diagram?
- Assemble a flow chart to show the events or steps.

These examples show a few qualities of reflective assessment, such as its dependence on questioning, reflecting, and various forms of communication. Teachers and students need only talk with each other about important questions, with or without pencils, dry boards, projectors, word processors, though these tools may facilitate reflective processes.

Asking questions and contemplating answers, both independently and collaboratively, are fundamental teaching and learning activities. Reflection, or forms of thinking synonymous with it, appear across cultures from ancient times. In the Old Testament, the psalmist reported meditating on the law of the Lord by talking to himself day and night (Psalm 1:2, The New King James Bible). The Greek sage, Aesop, told of an old woman who, chancing upon an empty wine bottle, recollected the once fragrant contents of the remaining dregs (Aesop, trans., 1992). In the Tao Teh Ching, the wise master, Lao Tzu, reminded the disciple that in order to cultivate the mind, one must "know how to dive in the hidden deeps" (trans., 1989, p. 17). In the Bhagavad Gita (2:41), the hero, Arjuna, was advised to contemplate one action at a time in order to avoid straying onto irresolute paths and innumerable

distractions.

These examples also show ways to focus student thinking on the purpose of the lesson in connection to previous, current, and subsequent learning activities. Some researchers associate this concept with alignment, or the accuracy with which elements of planning, instruction, and assessment work together to produce learning (Resnick, Rothman, Slattery, & Vranek, 2004). While most educators presume these elements work in concert with each other, there is evidence to suggest alignment is not always achieved, at the class level and at other levels of the education hierarchy (Browder, Spooner, Wakeman, Trela, & Baker, 2006; Parke & Lane, 2008; Pellegrino, 2006; Porter & Smithson, 2001; Tindal & Nolet, 1996). Gathering feedback from students about what they have learned, what they perceive as valuable, or what they believe is the purpose of a lesson or activity enables teachers and students alike to observe whether planning, instruction, and assessment are indeed working together to promote learning through alignment.

## Analysis of Quantitative Studies

A large body of empirical research exists regarding formative assessment strategies that occur during learning activities. A brief summary follows of some prominent studies that relate directly to reflective assessment. Since reflective assessment depends on characteristics of reflection, and also makes alignment of goals and activities more explicit, it is perhaps unsurprising that research shows positive effects of interventions indicative of reflective assessment on student achievement (Blank & Hewson, 2000; Bond & Ellis, 2013; Conner & Gunstone, 2004; Dignath & Büttner, 2008; Gulikers, Bastiaens, Kirschner, & Kester, 2006; Gustafson, 2002; Hartlep & Forsyth, 2000; Naglieri & Johnson, 2000; Schneider et al., 1986; Schunk, 1983; Wang, Haertel, & Walberg, 1993; White & Frederiksen, 1998). However, many of these findings are derived from studies examining various subjects, not the least of which include formative assessment and self-regulated learning, but also reflective thinking, critical thinking, questioning techniques, feedback, and strategy instruction.

176

---

One way to identify the effects of reflective assessment, in the context of limited or diversified research studies, is by applying best-evidence methodology. Similar to meta-analysis, the purpose is to reveal patterns, show relationships, and add to the cumulative knowledge of a particular field (Hunter & Schmidt, 2004; Slavin, Lake, Hanley, & Thurston, 2014). End goals of both meta-analysis and best-evidence summary is theory development or explanation of phenomena (Hunter & Schmidt, 2004; Slavin, Lake, Hanley, & Thurston, 2014). These techniques are especially important in the area of behavioral or social science investigations, given the limited number of studies in any one area, which often show conflicting results (Hunter & Schmidt, 2004). In addition, though less important as an underlying rationale for selecting best-evidence methodology, educators are becoming accustomed to reports of the parametric qualities of instructional practices, mostly from informative comparisons of effect sizes by Bloom (1984), Hattie and Timperley (2007), and Marzano, Pickering, and Pollock (2001).

Slavin, Lake, Hanley, and Thurston (2014) identify the following steps for conducting best-evidence syntheses: a) identify selection criteria for including or excluding studies, b) calculate average effect size across studies, c) weight effect sizes proportionally to the number of study participants to show results of practical or theoretical interest, and d) extend the description of results beyond quantities to encourage replication.

The analysis that follows is intended to provide educators with information on the effects of reflective assessment, or those reflective activities that integrate formative assessment based on self-regulation theory. The methodology adheres to steps for best-evidence synthesis outlined by Slavin, Lake,

Hanley, and Thurston (2014). The selection criterion includes studies with reflective assessment as the intervention. While this yields a small number of studies for inclusion, this approach avoids the “file drawer problem,” which Sheskin (2007, p. 1307) defines as omission of studies from the research record which show non-significant statistical results. In addition, the need to justify instructional practices, and their contribution to achievement – or in most cases, test achievement – is an increasingly important activity for educators. At the very least, an analysis justifying reflective thinking serves as an antidote to the current standards movement, which is justified through demands for accountability in the form of increasing test scores.

Questions used for guiding this best-evidence synthesis of the impact of reflective assessment on student learning include the following:

1. How does reflective activity affect performance on a content-specific posttest?
2. How does reflective activity affect performance on a content-specific retention test?
3. Does teacher feedback on reflective activities improve student performance on a content-specific posttest and retention test?

## Methodology

The data set included results from ten doctoral dissertations completed at one institution. Investigators conducted studies with the cooperation of teachers and school building administrators, where interventions were integrated as part of the assigned school curricula. All but one of the studies involved public school students. Moore (2010) sampled English speaking students from an international school in India. In total, study investigators worked with 1,251 students, grades 4 through 12, across a variety of content areas. Table 1 shows the author and context information for each study.

Table 1

<i>Study and Sample Characteristics</i>			
Author	<i>n</i>	Grade	Discipline
Bianchi	110	10	Science
Bond	141	5-6	Math
Denton	259	8	Social Studies
Edwards	54	9-10	Math
Evans	223	9	English
Johnson	65	4-6	Math
Kourilenko	85	9-12	French
Moore	73	4	Science
Shoop	134	9-12	Science
Zirkle	107	8	Geography
Total	1,251	na	na

Each study applied one or more reflective assessment strategies, as shown in Table 2. Eight of the studies used *I learned* statements in combination with *think aloud*, *talk about it*, *clear and unclear windows*, and *learning illustrated*. Two other studies used *journals* with *talk about it* or *read aloud*.

Reflective activities were deployed near the end of the lesson and required between five and ten minutes to complete. Participating teachers gathered student reflections as part of the intervention. However, six of the studies included teacher feedback on student reflections as part of the treatment, also shown in Table 2. An important part of feedback in some of the studies, when it was provided, included teachers identifying a few exemplary reflections from the previous day at the beginning of subsequent lessons, and in some cases, teachers leading brief discussions about the exemplars. Four of the studies involved teachers collecting reflections, but not providing feedback. Methods for providing feedback are shown in Table 2.

Table 2

*Independent Variables or Reflective Assessment Strategies*

Author	I Learned	Think A-loud	Talk About It	Journal	Read A-loud	Learning Illustrated	Clear and Unclear Windows
Bianchi	X						X
Bond	X	X					
Denton*†Δ				X	X		
Edwards*Δ	X	X					X
Evans*†Δ	X	X					
Johnson	X	X					
Kou-rilenko*†Δ	X		X				
Moore*			X	X			
Shoop*	X		X				
Zirkle	X					X	

\* Strategy deployed along with teacher feedback.

† Teacher shared exemplary student reflections at the beginning of each subsequent lesson.

Δ Teacher led discussion of exemplary reflections.

Each study included more than one classroom for assessing planned comparisons between treatment and comparison groups using analysis of variance (ANOVA). All studies included a control group, with the exception of Zirkle (2009). The dependent variable for each study was a content-specific test. Seven studies applied pre- and posttest design using equivalent forms of the content-specific test. Three of the studies applied a posttest only design. All but one study applied an equivalent form retention test, 6 to 12 weeks after posttest administration. Two studies included covariates. Four studies applied non-parametric calculations to mitigate non-normal distributions of posttest data.

Statistically significant results comparing treatment classroom and comparison classroom performance on content-specific posttests were reported in seven of the studies. Alternatively, treatment and comparison performance on posttests for three studies showed non-significant results. Levels of statistical significance for posttest comparisons are shown in Table 3.

Table 3

*Study Design Features*

Author	<i>n</i>	Treat- ment Compa- rison Control	Pretest and Posttest	Posttest Only	Re- tention Test	Covari- ate	Nonpa- rametric Tests	Posttest <i>p</i>
Bianchi	110	X	X		X			< .01
Bond*	95	X		X	X		X	< .00
Denton	187	X	X		X	X		< .01
Edwards	54	X	X			X		.42
Evans	163	X		X	X		X	< .00
Johnson	46	X		X	X		X	< .00
Kou- rilenko	59	X	X		X		X	.05
Moore	53	X	X		X			.19
Shoop	68	X	X		X			.63
Zirkle	107		X		X			.05

Note. \*Randomly assigned students to treatment, control, and comparison groups.

## Results

A fixed effects model was used for calculating a weighted mean effect size for posttest performance and retention test performance. According to Borenstein, Hedges, Higgins, and Rothstein (2009), fixed effects modeling is appropriate when studies use similar methodology and examine similar variables. To calculate weighted mean effects, Coehn's *d* effect sizes were calculated for each study and aligned with the number of participants as shown in Table 4 for posttest results, and in Table 5 for retention test results. According to Vogt (2005), individual study effect sizes, such as Coehn's *d*, show the estimated amount of variance on the dependent variable which may be attributed to an independent variable. Fixed effects modeling uses effect size estimates to calculate a weighted average based on individual case performance, rather than random effects modeling, which uses individual studies as the unit of analysis (Borenstein et al., 2009).

Table 4

*Posttest Effect Size Statistics*



Author	<i>n</i>	Cohen's <i>d</i>	<i>d * n</i>
Bianchi	110	.628	69.08
Bond†	95	.495	47.03
Denton*	187	.032	5.98
Edwards*	54	-.23	-12.42
Evans†	163	.69	112.47
Johnson†	46	-.19	-8.74
Kourilenko†	59	.395	23.31
Moore	53	.67	35.51
Shoop	68	-.34	-23.12
Zirkle	107	.15	16.05
Total	942		265.14

\* Included a covariate and conducted ANCOVA calculations.

† Included nonparametric calculations because data sets violated assumptions of normality.

Table 5  
*Retention Test Effect Size Statistics*

Author	<i>n</i>	Cohen's <i>d</i>	<i>d * n</i>
Bianchi	110	.31	34.10
Bond	92	.52	47.84
Denton	174	-.17	-29.58
Evans	158	.95	150.1
Johnson	46	-.22	-10.12
Kourilenko	59	.30	17.70
Moore	53	.43	22.79
Shoop	68	.06	4.08
Zirkle	107	.17	18.19
Total	867		255.10

Weighted mean effect sizes are calculated by multiplying the effect size of each individual study, *d*,

by the sample size for each study,  $n$ , as shown in Table 4 and 5. Results are then multiplied and the products are summed and divided by the total sample from all studies. The weighted mean effect size for the posttest was .28, while the weighted mean effect size for the retention test was .29.

Analysis of studies which included teacher feedback as part of the intervention showed similar effect sizes on posttest and retention tests. One exception was studies incorporating teacher feedback on reflections along with teacher led discussion of exemplary reflections, which showed a mean weighted effect size of .33 for retention tests.

Additional calculations were made which excluded studies by Shoop (2006) and Edwards (2008) because of high rates of student mortality, above 20%, and inconsistent deployment of intervention activities, according to limitations reported by each author. Revised calculations showed a weighted mean effect size of .37 for the posttest and .31 for the retention test.

## Conclusion

Use of reflective activities, specifically reflective assessment, showed positive effects on student learning. Weighted mean effect sizes for posttest and retention tests ranged from .28 to .37. For comparison, Hattie, Biggs, and Purdie (1996) report corrective feedback shows an effect size of .65, homework .43, and ability grouping .18. Also, in their early analysis of the effects of formative assessment, Black and Wiliam (1998) report effect sizes between .40 and .70. However, also according to Hattie, Biggs, and Purdie, the average effect size of an educational intervention is .40. Also for comparison, according to Cohen (1969), an effect size of .20 is small, .50 is medium, and .80 is large. However, Glass, McGaw, and Smith (1981) caution that the magnitude of effect size should be judged in comparison to similar interventions seeking to produce the same results. Some interventions similar to reflective assessment include higher-order questions, feedback with goals, and questioning, which show effect sizes of .30, .42, and .41, respectively (Bloom, 1984; Hattie & Timperley, 2007).

Lastly, an important point for accurate interpretation of results is the brief amount of time needed to engage students in reflective activities, which ranged from five to ten minutes. Minimal expenditure of time, in comparison to meaningful gains in student achievement, provide helpful context for judging small effect sizes as encouraging and worth further investigation.

## References

- Aesop (1992): *Aesop's fables* (J. Zipes, Trans.). New York, NY: Penguin.
- Bianchi, G.A. (2007): *Effects of metacognitive instruction on the academic achievement of students in the secondary sciences* (Unpublished doctoral dissertation). Seattle Pacific University, Seattle, WA.
- Black, P. & Wiliam, D. (1998): Inside the black box. In: *Phi Delta Kappan*, 80 (2), pp. 139-148. Retrieved from: <http://www.pdkintl.org/utilities/archives.htm>
- Black, P. & Wiliam, D. (2009): Developing a theory of formative assessment. In: *Educational Assessment, Evaluation and Accountability*, 21, pp. 5-31. (doi 10.1007/s11092-008-9068-5)
- Blank, L. M. & Hewson, P. W. (2000): A metacognitive learning cycle: A better warranty for student understanding? In: *Science Education*, 84 (4), pp. 486-506.
- Bloom, B. (1984): The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. In: *Educational Researcher*, 13 (6), pp. 4-16.
- Bloom, B. S., Hastings, J. T. & Madaus, G. F. (1971): *Handbook on formative and summative evaluation of student learning*. New York, NY: McGraw-Hill.

- Bond, J. B. & Ellis, A. K. (2013): The effects of metacognitive reflective assessment on the achievement of fifth and sixth grade students. In: *School Science and Mathematics*, 113 (5), pp. 227-234. (doi:10.1111/ssm.12021)
- Bond, J. B. (2003): The effects of reflective assessment on student achievement (Unpublished doctoral dissertation). Seattle Pacific University, Seattle, WA. Retrieved from <http://digitalcommons.spu.edu/etd/1/>
- Borenstein, M., Hedges, L., Higgins, P. & Rothstein, H. (2009): *Introduction to meta-analysis*. Hoboken, NJ: John Wiley & Sons.
- Browder, D., Spooner, F., Wakeman, S., Trela, K. & Baker, J. (2006): Aligning instruction with academic content standards: Finding the link. In: *Research and Practice for Persons with Severe Disabilities*, 31, pp. 309-321.
- Cohen, J. (1969): *Statistical power analysis for the behavioral sciences*. New York, NY: Academic Press.
- Conner, L. & Gunstone, R. (2004): Conscious knowledge of learning: Accessing learning strategies in a final year high school biology class. In: *International Journal of Science Education*, 26 (12), pp. 1427-1443. (doi: 10.1080/0950069042000177271)
- Denton, D.W. (2010): The effects of reflective thinking on middle school students' academic achievement and perceptions of related instructional practices: A mixed methods study (Unpublished doctoral dissertation). Seattle Pacific University, Seattle, WA. Retrieved from <http://digitalcommons.spu.edu/etd/13/>
- Dignath, C. & Büttner, G. (2008): Components of fostering self-regulated learning among students: A meta-analysis on intervention studies at primary and secondary school level. In: *Metacognition and Learning*, 3 (3), pp. 231-264. (doi: 10.007/s11409-008-9029-x)
- Edwards, T. G. (2008): Reflective assessment and mathematics achievement by secondary at-risk students in an alternative secondary school setting (Unpublished doctoral dissertation). Seattle Pacific University, Seattle, WA.
- Ellis, A. K. (2001, 2010): *Teaching, learning, and assessment together*. Larchmont, NY: Eye on Education.
- Evans, L. (2009): Reflective assessment and student achievement in high school English (Unpublished doctoral dissertation). Seattle Pacific University, Seattle, WA.
- Flavell, J. H. (1976): Metacognitive aspects of problem solving. In: L. B. Resnick (Ed.): *The nature of intelligence*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers, pp. 231-236.
- Glass, G. V., McGaw, B. & Smith, M. L. (1981): *Meta-analysis in social research*. London: Sage.
- Gulikers, J., Bastiaens, T. J., Kirschner, P. A. & Kester, L. (2006): Relations between student perceptions of assessment authenticity, study approaches and learning outcome. In: *Studies in Educational Evaluation*, 32 (4), pp. 381-400. (doi: 10.1016/j.stueduc.2006.10.003)
- Gustafson, K., & Bennett, W. (2002): Issues and difficulties in promoting learner reflection: Results from a three-year study. Retrieved from <http://www.stormingmedia.us/61/6162/A616274.html>
- Hartlep, K. H., & Forsyth, G. A. (2000): The effect of self-reference on learning and retention. In: *Teaching of Psychology*, 27, pp. 269-271. (doi: 10.1207/S15328023TOP2704\_05)
- Hattie, J. & Timperley, N. (2007): The power of feedback. In: *Review of Educational Research*, 77 (1), pp. 81-112. (doi: 10.3102/003465430298487)
- Hattie, J., Biggs, J. & Purdie, N. (1996): Effects of learning skills interventions on student learning: A meta-analysis. In: *Review of Educational Research*, 66 (2), pp. 99-136.
- Hunter, J. E. & Schmidt, F. L. (2004): *Methods of meta-analysis*. Thousand Oaks, CA: Sage.

- Johnson, L. I. (2004): The effects of reflective assessment on intermediate grade student achievement in mathematics. (Unpublished doctoral dissertation). Seattle Pacific University, Seattle, WA.
- Knowles, M. (1990): *The adult learner: A neglected species* (4<sup>th</sup> ed.). Houston, TX: Gulf.
- Kourilenko, I. N. (2013): Reflective assessment, feedback, and student achievement in foreign language studies: A mixed methods study (Unpublished doctoral dissertation). Seattle Pacific University, Seattle, WA.
- Marzano, R. J., Pickering, D. J. & Pollock, J. E. (2001): *Classroom instruction that works*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Medina, J (2008): *Brain rules: 12 principles for surviving and thriving at work, home, and school*. Seattle, WA: Pear Press.
- Moore, C. R. (2010): Mediated reflection and science achievement of fourth grade students in a highly diverse international school (Unpublished doctoral dissertation). Seattle Pacific University, Seattle, WA.
- Naglieri, J. A. & Johnson, D. (2000): Effectiveness of a cognitive strategy intervention in improving arithmetic computation based on the PASS theory. In: *Journal of Learning Disabilities*, 33 (6), pp. 591-598. (doi: 10.1177/002221940003300607)
- Parke, C. & Lane, S. (2008): Examining alignment between state performance assessment and mathematics classroom activities. In: *Journal of Educational Research*, 101, pp. 132-147.
- Porter, A. & Smithson, J. (2001): *Defining, developing, and using curriculum indicators* (Report No. CPRE-RR-Ser-048). Philadelphia: Consortium for Policy Research in Education. Retrieved from <http://www.cpre.org/Publications/rr48.pdf>
- Pressley, M. & Harris, K. (1990): What we really know about strategy instruction. In: *Educational Leadership*, 48 (1), pp. 31-34.
- Resnick, L., Rothman, R., Slattery, J. & Vranek, J. (2004): Benchmarking and alignment of standards and testing. In: *Educational Assessment*, 9 (1-2), pp. 1-27.
- Schneider, W., Borkowski, J.G., Kurtz, B. & Kerwin, K. (1986): Metamemory and motivation: A comparison of strategy use and performance in German and American children. In: *Journal of Cross-Cultural Psychology*, 17 (3), pp. 315-336. (doi: 10.1177/0022002186017003005)
- Schunk, D.H. (1983): Progress self-monitoring: Effects on children's self-efficacy and achievement. In: *Journal of Experimental Education*, 51 (2), pp. 89-93. Retrieved from <http://www.jstor.org/stable/20151486>
- Sheskin, D. J. (2007): *Handbook of parametric and nonparametric statistical procedures*. New York, NY: Chapman & Hall.
- Shoop, K. A. (2006): Self-reflection, gender, and science achievement. (Unpublished doctoral dissertation). Seattle Pacific University, Seattle, WA.
- Slavin, R. E., Lake, C., Hanley, P. & Thurston, A. (2014): Experimental evaluations of elementary science programs: A best-evidence synthesis. In: *Journal of Research in Science Teaching*, 51 (7), pp. 870-901.
- Tindal, G. & Nolet, V. (1996): Serving students in middle school content classes: A heuristic study of critical variables linking instruction and assessment. In: *Journal of Special Education*, 29, pp. 414-432.
- Tzu, L. (1989): *Tao teh ching* (J. Wu Trans.). Boston, MA: Shambhala.
- Vogt, W.P. (2005): *Dictionary of statistics and methodology: A nontechnical guide for the social sciences* (3rd ed.). Thousand Oaks, CA: Sage Publications.

- Wang, M. C., Haertel, G. D. & Walberg, H. J. (1993): Toward a knowledge base for school learning. In: *Review of Educational Research*, 63 (3), pp. 249-294. (doi: 10.2307/1170546)
- White, B. C. & Frederiksen, J. (1998): Inquiry, modeling, and metacognition: Making science accessible to all students. In: *Cognition and Instruction*, 16 (1), pp. 3-118. (doi:10.1207/s1532690xci1601\_2)
- Winne, P. (2005): A perspective on state-of-the-art research on self-regulated learning. In: *Instructional Science*, 33, pp. 559-565. (doi:10.1007/s11251-005-1280-9)
- Zimmerman, B. J. (1989): A social cognitive view of self-regulated academic learning. In: *Journal of Educational Psychology*, 81 (3), pp. 329-339.
- Zimmerman, B. J. (2002): Becoming a self-regulated learner: An overview. In: *Theory into Practice*, 41 (2), pp. 64-72.
- Zirkle, D.M. (2009): Long-term potentiation principles to form an optimal repetition schedule. Unpublished doctoral dissertation (Unpublished doctoral dissertation). Seattle Pacific University, Seattle, WA.

## **About the Authors**

**Prof. Dr. John B. Bond:** Associate Professor of Educational Administration and Supervision, Educational Leadership, Seattle Pacific University (USA). Contact: bondj@spu.edu

**Prof. Dr. David W. Denton:** Assistant Professor of Education, School of Education/Graduate-Programs at Seattle Pacific University (USA). Contact: dentod@spu.edu

**Prof. Dr. Arthur K. Ellis:** Professor of Education, Director of the Center for Global Curriculum Studies at Seattle Pacific University (USA), Corresponding Professor, University of the Russian Academy of Education, Moscow. Contact: aellis@spu.edu

