Review of Educational Research
December 2014, Vol. 84, No. 4, pp. 609–639
DOI: 10.3102/0034654314540007

On the Effectiveness of Supplemental Instruction: A Systematic Review of Supplemental Instruction and Peer-Assisted Study Sessions Literature Between 2001 and 2010

Phillip Dawson
Monash University

Jacques van der Meer
University of Otago

Jane Skalicky
University of Tasmania

Kym Cowley
University of Newcastle

Supplemental instruction (SI)—variously known as peer-assisted learning, peer-assisted study sessions, and other names—is a type of academic support intervention popular in higher education. In SI sessions, a senior student facilitates peer learning between undergraduates studying a high-risk course. This article presents a systematic review of the literature between 2001 and 2010 regarding the effectiveness of SI. Twenty-nine studies met the inclusion criteria. Due to methodological heterogeneity and lack of consistency defining the SI treatment, qualitative synthesis methods were applied. For seven included studies, however, an effect size of SI participation on final grades was calculated, ranging from d = 0.29 to d = 0.60. The findings of the review are consistent with claims validated by the U.S. Department of Education in the 1990s that participation in SI is correlated with higher mean grades, lower failure and withdrawal rates, and higher retention and graduation rates.

Keywords: supplemental instruction, PASS, peer learning, effectiveness, systematic review

Supplemental instruction (SI) is an academic support program that employs successful later-year tertiary students to facilitate peer-learning sessions mostly attached to high-risk courses. Originating at the University of Missouri-Kansas City (UMKC) in 1973, SI is currently offered internationally to hundreds of thousands of students each year (Arendale, 2002), at a cost we estimate to be in the tens of millions of dollars. Staff from more than 1,500 tertiary institutions across 29 countries have been trained to implement SI (Martin, 2009). To justify this substantial investment, members of the practitioner community are often challenged to provide research evidence on the efficacy of their SI programs and the SI model itself. This article documents a systematic review of published, peer-reviewed research from 2001 to 2010 into the effectiveness of SI.

SI is often attached to specific *high-risk courses*, a term that is intentionally left open to interpretation (Martin & Arendale, 1993). High-risk courses may include the following characteristics:

> Large amounts of weekly readings from both difficult textbooks and secondary library reference works, infrequent examinations that focus on higher cognitive levels of Bloom's taxonomy, voluntary and unrecorded class attendance, and large classes in which each student has little opportunity for interaction with the professor or the other students. (Arendale, 1994, pp. 11–12)

Common high-risk courses include first-year science, technology, engineering, and mathematics (STEM) subjects, although SI has been applied across a broad cross-section of disciplines. The program integrates academic skills with course content in a series of peer-facilitated sessions that are voluntarily attended by students enrolled in these courses.

Each SI session is attended by a group of students enrolled in the target course and is facilitated by an SI Leader. Typically, leaders are academically successful students with good interpersonal skills who recently completed the course and achieved a good grade. They are recruited, trained, and supervised by an SI Supervisor. The leader is not a tutor or teaching assistant; their role is not to introduce new content or to reteach lecture material. Instead, the leader is responsible for facilitating discussion around course content and related study skills, and for preparing learning activities such as worksheets, group work, problem-solving exercises, or mock exams for their students. The students who attend SI sessions are responsible for teaching each other the course content and for working together to solve problems. Leaders typically act as model students by attending lectures, taking notes, reading the materials assigned to the students, and demonstrating effective study skills.

When compared with other academic support interventions, SI programs are somewhat homogeneous, internationally. This similarity may be due to the activities of the International Center for Supplemental Instruction, which operates out of UMKC and maintains a network of Certified SI Trainers and National Centers for SI, which train and accredit SI supervisors. Internationally, every region is supported by at least one National Center/Trainer, including the United States of America, Canada, the United Kingdom, Europe, South Africa, and Australasia. Whereas the undergraduate student mentoring literature reveals substantial diversity in what we call mentoring for students (Crisp & Cruz, 2009), efforts have been made to build consistency in implementations of SI through training, manuals, conferences, and an e-mail list.

SI supervisors and leaders usually share a tacit understanding that SI is an effective academic support program, however effective is understood. In-house program evaluations are often conducted comparing the grades of SI participants with the grades of nonparticipants, and usually those who participate in SI do better (e.g., International Center for Supplemental Instruction, 2003). These notions were formalized and exposed to greater scrutiny by the U.S. Department of Education in 1992 (Martin & Arendale, 1993; U.S. Department of Education, 1995). Their validation of SI has been cited hundreds of times in the scholarly literature and contains the following specific claims:

1. Students participating in SI within the targeted high-risk courses earn higher mean final course grades than students who do not participate in SI. This finding is still true when analyses control for ethnicity and prior academic achievement.
2. Despite ethnicity and prior academic achievement, students participating in SI within targeted high-risk courses succeed at a higher rate (withdraw at a lower rate and receive a lower percentage of [fail] final course grades) than those who do not participate in SI.
3. Students participating in SI persist at the institution (reenroll and graduate) at higher rates than students who do not participate in SI (Martin & Arendale, 1993).

Claims 1 and 2 are supported by data analyzed by the UMKC team from 49 U.S. institutions, representing 1,447 individual courses and an undisclosed number of students (Martin & Arendale, 1993). Three studies (N = 1,689; N = 349; N = 1,628) into student persistence and SI at UMKC were used for Claim 3. On the strength of these claims and their supporting evidence, SI was certified as an exemplary education program and was eligible for federal funding, which led to the proliferation of SI programs across Northern America in the 1990s; for a history of this period, see Arendale (2002).

The question of SI's effectiveness, however, was not completely answered by this research from the early 1990s. Subsequent research into the effectiveness of SI has presented methodological critiques of earlier work. McCarthy, Smuts, and Cosser (1997), for instance, criticized earlier attempts to assess the effectiveness of SI. Treatment of SI attendance as a binary variable (students either attended or did not attend) was described as simplistic, with McCarthy et al. preferring a discrete variable for the number of sessions attended. They also questioned the usefulness of preentry test scores as a proxy for self-selection, as they claimed such scores may not necessarily be correlated with success in tertiary study. The results of the McCarthy et al. case study do not, however, contradict those described by Martin and Arendale (1993).

More than a decade after McCarthy et al.'s (1997) critique of the literature, there is still limited research that controls for motivation, as distinct from prior academic achievement, when evaluating the effectiveness of SI. Some studies into the effectiveness of SI (e.g., Bowles & Jones, 2003–2004b; Hensen & Shelley, 2003; Hodges, Dochen, & Joy, 2001) do not cite the work of McCarthy et al. (1997) and appear to be unaware of the issues they raise. Bowles, McCoy, and Bates (2008), for example, who did cite the work of McCarthy et al. (albeit incorrectly referenced), still viewed motivation as a function of prior academic achievement, and they still treated SI attendance as a binary variable.

Further complications arise in the selection of dependent variables in research evaluating SI. Much research investigates final course grades or retention, with an assumption that these are proxies for learning. However, Ashwin (2003) argued that this is not a safe assumption to make. Through a mixed-methods study of a non-SI peer-learning program, he found that although attending students were more likely to succeed in their courses, they adopted less "meaning orientated" (p. 169) approaches to their studies. The qualitative component of his study found that attendees developed an "increased awareness of the assessment demands of the course and that these students had become more strategically orientated in their approach to studying" (Ashwin, 2003, p. 159). Learning and grades are hopefully related, but it is naive to assume they are the same thing; a recent meta-analysis by Richardson, Abraham, and Bond (2012) partially confirms Ashwin's suspicions that strategic approaches to learning might yield greater results than deep approaches. Ashwin made the methodological argument that diverse mixed-methods evaluations of peer-learning programs are necessary to understand their true effectiveness.

Alongside Ashwin's (2003) critique of measures of effectiveness, and McCarthy et al.'s (1997) critique of measures for self-selection, is Kochenour et al.'s (1997, p. 578) critical review of the literature, which finds that "much is anecdotal, is based on small or non-representative samples, or does not adequately consider student ability as a possible explanation for the apparent 'effect' of SI." Questions of self-selection bias were raised in relation to motivation, achievement, and ability variables in research on SI in the 1990s; however, the research rarely addresses all three. Where motivation is addressed, it is sometimes simplistic questioning of students about their self-identified likelihood to attend SI (e.g., Arendale, 1997), and other times, more sophisticated tools are used, such as Baker and Siryk's (1984) Academic Motivation Scale.

At the turn of the millennium, research into SI was patchy and not universally methodologically strong. In addition, review studies typically employed narrative methods and nonsystematic meta-analyses of unpublished results (e.g., Arendale, 1997); they also were conducted by organizations that could be considered to have a vested financial interest in the success of SI. During the 2000s, however, SI enjoyed a resurgence in international popularity and was accompanied by a new body of research from Europe, Australasia, and South Africa, as well as Northern America. This systematic review into the effectiveness of SI and related programs was conducted against a backdrop of increasing participation in higher education in Australasia (Bradley, Noonan, Nugent, & Scales, 2008). There is a need for interventions to support greater bodies of more diverse students at financially constrained tertiary institutions (e.g., as a consequence of the 2008 financial crisis) that increasingly demand greater cost/benefit analyses of any resources allocated to nonessential provisions. Considering publication of review studies covering the period before 2001, the resurgence of SI since the late 1990s, and resource constraints, this study is limited to the new body of research from 2001 to 2010.

Method

Systematic review methodology takes a structured approach to gathering, assessing, and synthesizing literature relating to a particular question (Pawson, 2006). The research question for this systematic review was, "What is the effectiveness of Supplemental Instruction for the attending students?" The term *effectiveness* was defined broadly to enable inclusion of any form of empirical evidence, qualitative or quantitative, gathered through any combination of research methodologies, as long as it related to the students who attended the sessions.

Search Strategy

SI operates under a variety of synonyms: in Northern American contexts, it often operates under its original name, SI, but in the Australasian context, it is often called Peer Assisted Study Sessions (PASS), and in the United Kingdom, it is sometimes referred to as Peer Assisted Learning. A list of synonyms for SI was used as keywords for the search strategy, which was developed in consultation with the international practitioner communities established by the International Center for Supplemental Instruction. National Centers for SI across the world were asked for synonyms for Supplemental Instruction; these bodies were considered the expert panel of opinions by which the search terms ought to be constructed. The following synonyms for SI were provided by the expert panel: Extending the Class, Facilitated Study Groups, Meet-Up, Peer Assisted Learning, Peer Assisted Study Sessions, Peer Led Undergraduate Study, Peers Assisting Student Success, Review With a Peer, Structured Study Sessions, Study Group Learning, Supplemental Instruction, Supplemental Learning, and Supported Learning Groups. Searches on title and abstract of these terms were performed on the following databases: PsycINFO, ERIC, and Education Research Complete.

These databases were selected as they were the most commonly used databases for recent articles published in the *Review of Educational Research*. In addition, the search was also performed on *Google Scholar*, although we acknowledge that there is some debate around the role of Scholar in systematic reviews (Gehanno, Rollin, & Darmoni, 2013; Giustini & Boulos, 2013). The keywords *Learning Groups, Strategic Learning*, and *Study Groups* were removed from the strategy as these produced only irrelevant results. A manual search of a publicly available annotated bibliography maintained by the International Center for Supplemental Instruction was also included (SI Staff from UMKC, 2010).

Inclusion/Exclusion Criteria

For the initial search, research assistants were instructed to exclude articles if they were obviously not about SI or a related program. What constituted SI, or an SI-like program, was determined by the authors in consultation with the expert panel. The authors (who are all experienced SI practitioners and researchers) then conducted an in-depth inclusion/exclusion exercise of the remaining set of articles, with each of them being considered initially by two authors against the following criteria. Articles needed to discuss outcomes for students attending SI sessions to be included. Articles that solely discussed outcomes for SI leaders who run the sessions were not included. The SI sessions needed to be face-to-face for an article to be included. Although acknowledging there is a diversity of online SI approaches, they were not of interest for the purpose of this study. Sessions needed to be regular; articles discussing one-off workshops were not included. Articles needed to consider SI sessions involving groups of students; where the SI support was primarily one-to-one, the study was not included. Included articles needed to consider SI sessions which were run by a student leader (or mentor, facilitator, etc.) who was not concurrently enrolled in the course. Where groups were facilitated by an academic or peer-facilitated by a *current* student of the course, the article was not included.

Articles needed to consider SI sessions which were attached to a specific post-secondary program or course. Only articles published in a peer-reviewed journal or the proceedings of a peer-reviewed conference were considered. Where the full text of an article could not be obtained through our collective institutional libraries, we contacted the authors of the study to request a copy; articles were excluded if we could not obtain a full-text copy. Finally, only one copy of each study was included. If the results of an included study was published multiple times (such as a conference paper republished in a journal with no changes to method or results), then the later publication was included but not the earlier one.

Data Extraction

Each included article was read by all four researchers, and structured data were extracted from each one. In addition to basic details about each study, data extraction focused on (a) the institutional and course context of each study; (b) characteristics of the participants; (c) the research approach taken; (d) any deviations from or augmentations to SI; (e) claims about the effectiveness of SI; and (f) evidence for those claims. Formal method quality analysis was not undertaken due to the methodological diversity of the set of articles. Instead, a qualitative discussion of issues of method quality is included when it is relevant to reporting about each article.

Synthesis

An initial reading of the articles revealed that meaningful quantitative synthesis (such as meta-analysis) would not be possible due to methodological heterogeneity, poor method quality, and insufficient description of method; additionally, the data set contained many qualitative results. Our synthesis, therefore, was qualitative and most similar to a thematic analysis approach (Bearman & Dawson, 2013; Dixon-Woods, Agarwal, Jones, Young, & Sutton, 2005). Where possible, an effect size was calculated for the relationship between SI and final grade. However, we caution readers to consider these in the context of the method quality of each study and its definition of SI treatment. Inclusion/exclusion and data extraction processes were piloted and refined iteratively: first on one article, then on 10% of the articles, before being implemented on the entire set. After each piloting, the researchers met face-to-face and refined procedures.

Where there was disagreement about inclusion/exclusion or data extraction, the research team again met face-to-face and discussed until a consensus was reached. The researchers are also authors of other research studies that were considered for inclusion in this work, and potential conflicts of interest were avoided by excluding any researcher from decisions about their own articles.

Results and Discussion

The initial search strategy yielded 596 results from databases and 819 results from Google Scholar. Research assistants selected 98 of these articles based on title and abstract, that is, they appeared to fit many of the basic inclusion criteria. Five additional articles were identified from the UMKC SI Bibliography resulting in a total of 103 articles based on title and abstract. The authors then read each article and included 29 of them in the final set. Some basic details about each study are presented in Table 1.

Definition of SI

To understand the effectiveness of SI for students, we first looked for a consistent definition of what the SI treatment is. Unfortunately, SI was not consistently defined in the included articles, but the studies frequently cited key works in the literature (i.e., Blanc, DeBuhr, & Martin, 1983; Martin & Arendale, 1993; Martin, Arendale, & Blanc, 1997). The use of the term *supplemental instruction* was not necessary for articles to be included; however, the term appeared in every included article at some point. It is noted that social interventions (like SI) are "leaky and prone to be borrowed" (Pawson, 2006, p. 32): What is called SI in one circumstance is not necessarily the same intervention as SI in another circumstance. To assume that a program is SI because it is called SI would be "label naivety" (Øvretveit & Gustafson, 2002, as cited in Pawson, 2006, p. 33) and would conflate diverse programs that might have nothing in common.

The absence of an unambiguous working definition of what an SI session is, and what happens in one, is relatively consistent with literature about other academic support interventions, such as Crisp and Cruz's (2009) review of undergraduate mentoring literature, which found more than 50 definitions of mentoring, but few discussions of what mentoring meant in practice. In our review, few articles specified what happens in the actual SI sessions. Rath, Peterfreund, Xenos, Bayliss, and Carnal (2007, p. 207) provided a description of a sample SI session:

> Typical activities included guided discussions with extensive class participation (often following small group work), worksheets that were completed both individually and in groups, peer instruction, preparation of study resources, kinesthetic and visual modeling of problems, practice tests, and trivia-style games. Particular emphasis was placed on the concepts, content, and vocabulary from the lecture, but before lab exams some time was spent reviewing methods, data analysis, and the interpretation and principles underlying observed outcomes of various laboratory experiments.

When claims of what happened in the SI sessions were given, like Rath et al.'s (2007), they were usually not accompanied by evidence and may be aspirations or expectations rather than actual observations. The SI model as described by the SI Supervisor manual (SI Staff from UMKC, 2005) includes substantial observation

TABLE 1

Overview of 29 included articles

Study	Field of study	Method	Study participants	Self-selection	Effectiveness
Bowles and Jones (2003–2004a)	24 diverse disciplines	Quantitative: OLS vs. simultaneous limited dependent variable model	3,645	Precourse GPA; pretertiary achievement	Course grade
Bowles and Jones (2003–2004b)	24 diverse disciplines	Quantitative: Bivariate probit model	3,646	No	Retention
Bowles, McCoy, and Bates (2008)	Not stated	Quantitative: Two equation treatment effects	3,905	Pretertiary achievement	Graduation
Bronstein (2008)	Physical chemistry	Mixed-methods: Case study	Not stated	Pretertiary achievement	Course grade; pass/fail; retention; well-being
Cheng and Walters (2009)	College algebra and probability; Precalculus 1	Quantitative: Observational	534	Pretertiary achievement	Pass/fail
Congos and Mack (2005)	Chemistry	Mixed-methods: Comparison between groups; anecdotes	"Thousands"	Pretertiary achievement	Course grade; pass/fail
Court and Molesworth (2008)	Creative media production	Quantitative: Survey	34	Pretertiary achievement	Skills development; social; uncertainty reduction
Dancer, Morrison, and Smith (2007)	Econometrics	Quantitative: Econometric models and descriptive statistics	890	Precourse GPA; pretertiary achievement	Course grade
Dobbie and Joyce (2008)	Accounting	Qualitative: Thematic analysis of focus groups	12	N/A	Academic skills; connectedness (student–student); capability (academic competence)
Fayowski and MacMillan (2008)	First year calculus for nonmajors	Quantitative: Quasi-experimental	869	Precourse GPA; pre-tertiary achievement; comparison with prior offering that was not provided SI	Course grade; pass/fail
Gattis (2002)	Chemistry	Quantitative comparison; ANCOVA	142	Pretertiary GPA; preadmission test; motivational control: Expressed desire to attend	Course grade

(continued)

TABLE 1 (continued)

Study	Field of study	Method	Study participants	Self-selection	Effectiveness
Hafer (2001)	English composition	Quantitative comparison	1,500	Final grades comparison with previous offering	Course pass/fail; withdrawal
Hensen and Shelley (2003)	Biology; chemistry; mathematics; physics	Quantitative: Comparison; ANOVA; ANCOVA; chi-square	7,339	Pretertiary achievement	Course grade; pass/fail
Hodges and White (2001)	History; mathematics	Quantitative: Experimental posttest-only control group	103	Pretertiary achievement	Mean semester GPA; comparison with tutoring
Hodges, Dochen, and Joy (2001)	History: United States	Quantitative: ANOVA and post hoc comparisons	432	Nonrandom assignment to groups; motivation scale	Course grade; pass/fail; postcourse GPA
Longfellow, May, Burke, and Marks-Maran (2008)	English composition	Mixed methods: Evaluative research study; survey; interview; descriptive statistics	216	Nonrandom assignment to groups; motivation scale	Aggregate assessment scores in a subsequent course
Mahdi (2006)	Engineering and technology	Mixed-methods: Quantitative comparison; qualitative	25	Pre-SI task mark	Course marks
Miller, Oldfield, and Bulmer (2004)	Chemistry; statistics	Mixed-methods	1,089 chemistry students; unclear number of statistics students	Pretertiary achievement	Final course marks
Moore and LeDee (2006)	Biology	Quantitative: Quasi-experimental	Not stated; we infer 1,592 (398/semester over 4 semesters)	Pretertiary achievement	Course grade; pass/fail
Ning and Downing (2010)	Business	Quantitative: SEM; univariate analysis	430	Pretertiary achievement	Academic skills; GPA
Ogden, Thompson, Russell, and Simons (2003)	Political science	Mixed-methods	248	Pretertiary achievement; precourse GPA	Course grade; postcourse GPA; retention
Parkinson (2009)	Chemistry; mathematics	Quantitative: Quasi-experimental; ANOVA	66	Pretertiary achievement	Course grade; course pass/fail; task mark

(continued)

TABLE 1 (continued)

Study	Field of study	Method	Study participants	Self-selection	Effectiveness
Peterfreund, Rath, Xenos, and Bayliss (2008)	Chemistry; biology; statistics; calculus	Quantitative	12,423	Pretertiary achievement; precourse GPA	Course grade; pass/fail; retention
Phelps and Evans (2006)	Mathematics (developmental)	Mixed-methods	13,800	A control group that was not offered SI	Course grade; pass/fail
Rath, Peterfreund, Xenos, Bayliss, and Carnal (2007)	Biology	Quantitative	2698	Pretertiary achievement; comparison with a group that was not offered SI	Course grade; pass/fail; retention
Smith, May, and Burke (2007)	Surveying	Qualitative: Case study	332	Pretertiary achievement; comparison with a group that was not offered SI	Academic skills; connectedness (student–student); capability (academic competence)
Stansbury (2001)	Chemistry	Quantitative: Comparison	15	Pre-SI task mark; interview	Course grade; pass/fail; withdrawal
van der Meer and Scott (2009)	Not stated	Mixed methods	345	Pre-SI task mark; interview	Academic skills; connectedness (student–student); capability (academic competence)
Wright, Wright, and Lamb (2002)	Mathematics (developmental)	Quantitative	Not stated	Pretertiary achievement	Course grade; pass/fail

Note. OLS = ordinary least squares; SI = Supplemental Instruction; GPA = grade point average; ANOVA = analysis of variance; ANCOVA = analysis of covariance; SEM = structural equation modeling.

by an SI supervisor of what actually goes on in SI sessions, and it is possible that these sorts of claims are supported by observations, but these are rarely treated as research data.

To further illustrate the diversity of operationalization of SI, the place and number of participants varied substantially. Some articles gave an indication of the number of students who were involved in SI sessions, ranging from small numbers like three, four, or five students per session (e.g., Fayowski & MacMillan, 2008; Ning & Downing, 2010; Parkinson, 2009) up to 20 students (e.g., Rath et al., 2007). As a collaborative, facilitated environment, the effectiveness of SI may be influenced by the number of people in the room, but this information is not consistently provided. Additionally, the location of the sessions was usually not specified, but when it was, it ranged from small classrooms (e.g., Smith, May, & Burke, 2007) to specialized laboratories (e.g., Mahdi, 2006).

Most studies compared SI participants with nonparticipants; however, the definition of these two groups varied considerably. Some studies also used multiple definitions of SI participation in the one study. The most common approach was to decide on a number of sessions to use as a minimum to count a student as an SI participant: 1 session (Bronstein, 2008; Congos & Mack, 2005; Hensen & Shelley, 2003; Hodges et al., 2001; Longfellow, May, Burke, & Marks-Maran, 2008; Mahdi, 2006; Miller, Oldfield, & Bulmer, 2004; Phelps & Evans, 2006; van der Meer & Scott, 2009), 3 sessions (Bowles & Jones, 2003–2004a, 2003–2004b; Bowles et al., 2008; Wright, Wright, & Lamb, 2002), 5 sessions (Fayowski & MacMillan, 2008), or 12 sessions (Stansbury, 2001). Decisions for the cutoff number of sessions were largely arbitrary and unsubstantiated; when a rationale was provided, it was usually that an effect was expected after that amount of SI attendance (e.g., Fayowski & MacMillan, 2008; Stansbury, 2001). Sometimes students were divided into groups of attendance, such as attending 0, 1 to 3, 4 to 6, 7 to 9, 10 to 12 sessions (Dancer, Morrison, & Smith, 2007) or other similar breakdowns of attendance groups (Gattis, 2002; Hodges & White, 2001; Longfellow et al., 2008; Stansbury, 2001). Some others conducted analyses that used the number of sessions as a discrete variable (Cheng & Walters, 2009; Dancer et al., 2007).

Two troubling issues arose in understanding the reviewed articles' definition of participation in the SI program. In three cases, the requirements to be classed as an SI student were never explicitly defined (Dobbie & Joyce, 2008; Hafer, 2001; Smith et al., 2007). In five other articles, however, a student was classed as an SI participant simply by enrolling in SI (Moore & LeDee, 2006; Ning & Downing, 2010; Ogden, Thompson, Russell, & Simons, 2003; Peterfreund, Rath, Xenos, & Bayliss, 2008; Rath et al., 2007). One concern is that researchers may be measuring the effect of intent to attend SI, rather than the effect of participation in SI.

Effectiveness

The discussion of the effectiveness of SI is to some extent framed by the three claims validated by the U.S. Department of Education, as discussed in the introduction. In summary, it was claimed that students' participation in SI results in higher mean marks, a higher percentage of students who pass the course, and an increased retention and graduation rate. It was asserted that these higher means were also achieved when considering ethnicity and academic achievement on entry. In addition to considering the three validated claims against the included studies, this review also considers whether research evidence was available for other claims of effectiveness of SI. With regard to the latter, there are a range of studies with additional effectiveness claims, such as developing academic skills and enhancing peer relationships.

In this section, an overview of the evidence will be provided for these claims. The most common measures of effectiveness that were reported—final course grades and course completion rates—are described first. These are followed by studies that controlled for a range of factors, including prior achievement (often as a proxy for ability), motivation, and college grade point average (GPA; as proxy for ability and/or motivation). Many effectiveness studies, however, do not, or do not adequately, control for possible explanatory variables. Lastly, effectiveness related to other characteristics and categories is considered.

Final Course Grades

By far the most common outcome of the effectiveness of SI was based on final course grades; 16 studies used this dependent variable. In many cases, effectiveness was assessed by comparing the course grades of those who participated in SI and those who did not. Most of these studies employed a quasi-experimental design: Students were not randomly assigned to either of these two groups but opted into or did not opt into SI. In general, most studies included significance tests for the differences between the means of the two groups and found that these differences were statistically significant, but the studies did not include effect sizes, or standard deviations (that would have allowed effect sizes to be calculated). Studies from Australia typically used marks out of 100 (e.g., Dancer et al., 2007; Miller et al., 2004), and studies from the United States typically used a 4-point GPA scale.

As an example of some of these typical studies, Dancer et al. (2007) reported on an Econometrics course with 890 students, 262 of whom were attending SI. The difference between the marks for the non-SI ($M = 58.9$) and SI participants ($M = 64.1$) was significant, $p < .001$, $d = 0.39$. Hensen and Shelley (2003) studied four science and mathematics courses and reported a significant difference ($p < .05$) between SI and non-SI participants, but did not provide exact significance levels. A study on the results of SI over a period of 13 years in STEM courses (Peterfreund et al., 2008) indicated significant differences ($p < .005$) for most of the courses. Although most SI programs operate in first-year courses only, Peterfreund et al.'s study also included some upper level courses. In the entry level courses, the authors noted that SI contributed to increased pass rates, and in higher level courses, SI contributed to increased numbers of As and Bs.

Fayowksi and MacMillan (2008) reported a nearly two-point difference (on a self-devised 12-point scale) of the final mean grade for a mathematics course between SI and non-SI participants, and this finding was statistically significant, $F = 26.8$, $p < .0005$. The authors argued that using a 12-point scale enhanced the sensitivity of the analysis, but the creation of the scale was rather arbitrary. The letter grades and part grades from D to A+ were assigned the numbers 3 to 12, F was assigned 2, and withdrawing from the course was assigned 1.

Some studies employed a less conventional approach to report on differences in final grades. In one study, the effect of SI was calculated by considering the change in scores between two different English reading/writing modules (Longfellow et al., 2008). In their context, it is typical, Longfellow et al. (2008) argued, for second semester scores to be lower than first semester scores. The scores of students who had attended SI sessions showed a significantly ($p < .05$) lower reduction of scores.

Some studies compared grades of the years that SI was not implemented with grades of the years when SI was implemented. Bronstein (2008), for example, compared the results of a chemistry course in 2001 and 2006. She reported that grades of the 2006 class, with SI implemented, were significantly higher. No data, however, are provided to substantiate this difference. Similarly, Congos and Mack (2005) compared grades before and after introduction of SI in two elementary chemistry courses. Inconsistent and incomplete presentation of the data (e.g., no standard deviations are included), however, makes it difficult to interpret their significance. Also, whereas the differences in mean final grades between the years prior to introduction of SI were small, after implementation there was much greater variability. For some semesters, it seemed that influences other than the SI intervention were at play, possibly a change of course material, instructor, or delivery. Congos and Mack did not discuss these results. These omissions make it difficult to assess the impact of that particular implementation of SI.

A number of studies sought to assess the impact of the number of SI sessions attended on final grades. Mahdi (2006), in a study with 25 engineering students, asserted improvement of students who attended four sessions or more. However, as insufficient data were provided, it was not possible to verify the statistical significance of this claim. In an Australian study related to a first-year chemistry course, Miller et al. (2004) reported a statistically significant ($p < .05$) difference between the results of students who attended no SI sessions ($M = 4.51$) and students who attended more than five sessions ($M = 5.29$), and they provided insufficient information to verify this claim. This study is further hampered by absence of information on the number of the 1,131 students who attended SI and the grading scale used. Nonetheless, these authors reported that the average GPA of the chemistry class improved from 4.34 in the 2 years prior to implementation to 4.63, 4.61, and 4.91 since implementation. Although the growth in mean scores suggests a positive impact of SI, as the classes from year to year are composed of different groups of students, some analysis in relation to self-selection would have strengthened the conclusions.

We found one study (Hodges & White, 2001) that compared SI with another form of learning support referred to as tutoring. The aim of this study was to assess the effectiveness of self-monitoring and verbal prompts on attendance at SI and tutorials of 103 high-risk students on conditional enrolment. Although the final GPAs for SI ($M = 2.35$) and non-SI participants ($M = 1.95$) were significantly different, achievement did not differ significantly for students who attended ($M = 2.23$) or did not attend tutorials ($M = 2.10$). Effect sizes were not published in this study, but we were able to calculate these from the data provided. Cohen's d was 0.51 for the SI comparison and 0.16 for the tutoring comparison.

Course Completion

Some studies used the pass and failure rates for courses as the outcome measure. Comparisons between the pass and failure groups were carried out with varying levels of details, however. For example, Congos and Mack (2005) provided only percentages of SI and non-SI with passes (As, Bs, and Cs) and failures (D, Fs, and withdrawals). However, Hensen and Shelley (2003) performed a chi-square test and reported that SI participants had significantly better pass rates at the $p < .05$ level. Peterfreund et al. (2008) reported that pass levels for SI participants in most of the courses reached significance at the $p < .05$ level, and some at the $p < .005$ level.

Comparison before and after introduction of SI also figured in this category of effectiveness reporting. These pre-/postdata could be seen as providing some degree of controlling for self-selection: Assuming similar bodies of students and course offerings, if students perform better in semesters where SI was offered, then perhaps SI contributed to this improvement. Congos and Mack (2005) conducted this type of study and reported on the pass and failure rates of two chemistry courses. They found that the number of Ds, Fs, and Ws for SI participants (33%) was much lower than for non-SI students (54%) in one of the two courses. However, the average proportion of DFW's for the four semesters before introduction of SI was 45%; if SI was the only factor, the DFW rates of the pre-SI (45%) and non-SI (54%) groups would be expected to be similar. Additionally, there was substantial variation in DFW rates for SI participants between semesters, ranging from 6% to 50%. It is possible that other factors may better explain intersemester variation than the offering of SI.

Other studies were more successful in making a case for the usefulness of comparing data before and after introduction of SI. Fayowksi and MacMillan (2008) looked at success to failure rate proportions for the three groups (SI, non-SI, and pretreatment). A chi-square test suggested that the SI group did significantly better ($p < .0005$) than the other groups. The proportions of students succeeding and failing in the three groups suggests that the difference between the non-SI group and the pretreatment group the year immediately before SI was introduced was not statistically significant: 53% passing versus 47% failing for the non-SI group and 47% passing versus 53% failing for the pretreatment group. They used this similarity as justification for combining the pre-SI and non-SI groups in their subsequent analyses.

This category of course completion studies also include those that considered the probability of students passing a course because of attendance in an SI program. Cheng and Walters (2009) studied the probability of successful course completion of 534 (of 816) students enrolled in two mathematics courses. They conducted a logistic regression, including 16 possible variables that might predict students' success. They concluded that students who had attended all SI sessions were 10 times more likely to succeed than students who attended no sessions at all. They cautioned, however, that over a third of the records had missing data and could not be included.

Assessment Tasks Performance

Three studies sought to identify improvements in particular course work assessments other than final grades. Bronstein (2008), for example, claimed that homework assignments and tests in a particular course showed improvement because of SI. No data, however, were provided for this claim. Miller et al. (2004) provided data on achievement in the full range of assessment tasks in a first-year chemistry course. They reported on students' performance in laboratory practica, multiple-choice question exams, and computer-generated tests in relation to students' SI attendance data. Students attending five or more sessions had higher scores than students not attending SI on all three tasks. However, not enough information was provided to evaluate these results fully. Parkinson (2009) reported on the introduction of SI in a first-year Biotechnology course in Ireland and provided achievement data for four successive course tasks. The results of the experimental ($n = 24$) and control group ($n = 43$) revealed differences with progressively higher significance levels from the first assessment task to the fourth one.

The Question of Self-Selection and Effectiveness

Is it really the SI intervention that explains the differential outcomes of students who do or do not participate in SI? This question has occupied many researchers seeking to establish the effectiveness of SI. One of the prominent factors that some researchers seek to control for is self-selection into the SI program based on motivation, as SI participation is (usually) voluntary. Some do so by administering a measure of motivation before students' participation or enrolment in SI. Others use students' GPA before participation in SI as a proxy for motivation, the assumption being that GPA reflects students' effort in achieving their academic results.

Another factor that researchers have tried to control for is academic ability, assuming that there is a chance that more academically able students are more likely to enroll in SI rather than less academically able students. Variables that are typically used to control for this are students' university (college) entry scores, for example, SAT or ACT scores in U.S. studies. Overall, it can be said that where studies did seek to account for these factors, it was established that if there were any effect at all, SI was the more important of the two variables in explaining the enhanced achievement of SI participants. Some caution has to be observed, however. Neither university entry scores nor pre-SI GPA scores can be assumed to definitively control for either motivation or ability. If these input variables reflect in any way either motivation or academic ability to some degree, there is also likely to be an interaction effect between them.

The majority of studies using entry scores were conducted in an American context. Some studies merely reported the entry scores for comparison; others controlled for these. Congos and Mack (2005), for example, provided the SAT scores of the SI and non-SI groups as proxy for ability/motivation. For most semesters that they reported on, the differences were not significant. Peterfreund et al. (2008) provided the SAT scores for the students in the various courses in which SI was introduced. Where there were significant differences, SI participants had lower scores. Interestingly, high school GPA differences were not significant in any of the cases. Fayowksi and MacMillan (2008) controlled for GPA prior to SI. The analysis of covariance results established that SI and not prior GPA explained the difference in achievement. As already indicated above, this well-designed study draws on comparison between 990 non-SI participants (including both pre-SI students and non-SI students) and 269 SI participants. Apart from appropriately reporting on statistical significance, they also reported a moderate effect size ($d = 0.5$; Cohen, 1988) that exceeds the 0.4 threshold that Hattie (2009) suggested for any educational intervention to be considered practically useful.

To establish the effect of SI on pass/failure rates while controlling for ability/motivation, Fayowski and MacMillan (2008) performed both a sequential logistic regression ($p < .0005$) and a Wald test. The latter test established not only the statistical significance of SI in terms of higher pass rates but also the practical significance of this result. The odds of a student who participated in SI (with a minimum attendance of at least five sessions) earning a passing grade was 2.7 greater than that for nonparticipants; one unit increase in prior GPA (on a 12-point scale) yielded a slightly higher odds ratio of 3. Hensen and Shelley (2003) compared the composite ACT scores for SI and non-SI participants with an analysis of variance. SI participants in biology, chemistry, and mathematics courses had significantly ($p < .05$) lower scores, whereas there was no difference between the SAT scores for SI and non-SI physics students. Controlling for this preentry characteristic suggested that SI participants did better ($p < .05$). Although they did not provide specific significance levels, Hensen and Shelley did provide details on standard deviations of the mean grades. Using the data provided, we were able to calculate effect sizes for the four courses. These ranged between $d = 0.23$ and $d = 0.29$, considered small by Cohen (1988).

One study (Moore & LeDee, 2006) used a variation of the ACT score, the AAR (ACT Aptitude Rating), which incorporates both the ACT score and high school percentile multiplied by two. They found that the AAR of SI students was significantly lower ($p < .01$) than non-SIs, but that there was no significant difference in course marks. Moore and LeDee studied the results of four semesters with an average of 37 students attending SI, and 361 not attending. They considered SI successful, because SI students started off with lower AAR scores. They also argued that it was not just SI attendance that contributed to the success of these students but also other academic behaviors such as lecture attendance and help-seeking behavior (optional help session and use of visit hours). They considered these behaviors to be related to motivation. Moore and LeDee analyzed these academic behaviors for SI participants who earned an A or B grade and concluded that these students displayed more of the behavior they considered desirable than SI students who earned a D or F. Surprisingly, although they were interested in academic behaviors, they did not report on SI attendance patterns. SI participation was merely considered as a binary variable.

Two studies used Australian entry scores. Miller et al. (2004) provided students' overall position (OP) ratings (from 1 = high to 25 = low) as the preentry academic achievement variable. For the group of students included in the research, the OP range was from 1 to 16. In considering the relationship between OP scores and the number of SI sessions attended by students in a first-year chemistry course, Miller et al. provided the OP score for the different attendance bands. These ranged from 5.3 (no sessions attended) to 3.9 for students who attended five or more sessions. They did not, however, control for this OP measure. For a first-year statistics course, they performed a multiple regression test, and claimed that there was no evidence of an interaction effect between OP scores and SI attendance (p = .819). The level of SI attendance did have an impact on all students, with students with better OP scores doing relatively better when they attended a similar number of SI sessions compared with students with worse OP scores. Conversely, students with a less favorable OP score who attended more sessions did better than students with a better OP score who attended fewer sessions. As already mentioned earlier, this study unfortunately did not provide enough data to verify all claims. Another Australian study (Dancer et al., 2007) considered the university admissions index, which is a percentile rank. They found very small differences between SI and non-SI participants (92nd vs. 91.8th), and the correlation between university admissions index and SI participation was very small r = .07. Dancer et al. (2007) concluded from this that brighter students did not necessarily attend SI in higher proportions.

Another variation on using university entry scores as a proxy for academic ability was a composite indicator to predict a student's performance. In one study (Ogden et al., 2003), a predicted GPA, or PGA, was calculated that included both SAT scores, high school marks and achievement prior to SI being offered. In this study, a further differentiation was made by entry status of students enrolled in the course in which SI was introduced: traditional entry students and conditional entry students (who needed learning and/or language support). They found no difference between the predicted GPA of the traditional, conditional, and non-SI students. Conditional SI students (n = 17), however, did better (p < .05) than conditional non-SI students (n = 66). Their course grades came up to the same level as traditional students who did not participate in SI (n = 139) but not to the same level as traditional students who did participate in SI (n = 26).

Gattis (2002) used an Admissions Index, which is a predictor of first-year results derived from high school results and admissions information. This Index was mapped against course grades and the number of SI sessions attended. The results of eight consecutive semesters of a STEM course where SI was introduced showed that the Admissions Index for students did not differ for SI participants and nonparticipants, but that higher grades were gained by students who attended more SI sessions. The 2,606 students who did not participate in SI had an average Admissions Index of 2.83 and course grade of 2.15, whereas the 100 students who attended between 9 and 13 sessions had an Admissions Index of 2.81 and course grade of 2.86; the 60 students who attended more than 13 sessions had and Admissions Index of 2.78 and course grade of 3.06.

Another way one could control for motivation/ability is by making SI compulsory so that self-selection into the SI program by more able and/or more motivated students could be eliminated. One study that met the selection criteria of the systematic review attempted to do this (Hodges et al., 2001). This study, carried out in a compulsory course in a first-year history program, created three groups: a voluntary SI group (n = 105), non-SI participants (n = 219), and a group of students who were mandated to attend SI sessions (n = 108). This program reported final mean grades of 2.49, 2.13, and 2.74, respectively, for the three groups. Post hoc analyses indicated statistically significant differences between non-SI and the voluntary SI group, p = .0136, and between non-SI and the mandated SI group, p = .0000. Unfortunately, no effect sizes or standard deviations were provided.

Failure rates for the three groups (DFW) were 19% for the voluntary SI, 41% for the non-SI, and 9% for the mandatory-SI groups.

These results seem to suggest that when self-selection bias is not present, a broad range of students did considerably better through SI participation than students who did not participate in SI, or voluntarily opted into SI. To assess independently the levels of motivation of the different groups, Hodges et al. (2001) asked students to complete an academic motivation scale self-assessment (Baker & Siryk, 1984). Analyses indicated that the voluntary SI participants were significantly more motivated than the mandatory SI group (p = .0002) and the mandatory SI group (p = .0443). However, comparison of SAT scores revealed no significant differences among the groups: 871 for the voluntary group, 904 for the non-SI group, and 895 for the mandatory group. These researchers also looked at high school ranking and reported that the rank score for the voluntary group was higher (p = .03) than the two other groups. These results might leave us with the conclusion that, although the mandatory group was less motivated and had lower preentry scores than the students who chose to participate in SI, they did considerably better than the non-SI participants. It does not, however, explain the difference between the voluntary and mandatory SI participants. Surprisingly, the researchers did not measure the effect of one salient factor, the effect of number of SI sessions attended. They reported that the mandatory group had to attend all 13 sessions, but this was not the case for the voluntary group. Few students in this latter group, they say, attended more than 10 sessions, however.

A further variation in controlling for self-selection involved considering students who wanted to attend SI but were prevented from doing so because of timetable clashes or other commitments as a control group. Gattis (2002) found that students who attended four or more sessions (n = 41) did better than those who chose not to attend (n = 48), with a mean grade of 2.86 versus 2.24, and also better than the control group 2.70 (n = 11). We calculated the effect size for these differences as d = 0.46 and d = 0.14, respectively. Gattis subsequently adjusted (least squared means) the grades for preentry characteristics as measured by the Admissions Index. This adjustment resulted in modified grades of 2.96 for SI participants, 2.04 for nonparticipants, and 2.48 for the control group. Effect sizes could not be calculated as no standard deviations were provided for this.

In one study, Parkinson (2009) created an experimental and control group from a first-year group of Biotechnology students. All incoming Biotechnology students (N = 67) had volunteered for SI, but only 24, who were deemed to be representative of the composition in terms of preentry characteristics of the whole class, were assigned to the group who would receive the SI treatment. Differently than other SI programs, SI sessions in this study were connected to the whole program of study rather than to individual courses. Performance on four assessment tasks over the whole semester as well as three end-of-semester examinations showed significant differences between the experimental and control group ($.001 < p < 0.006$), in contrast with no significant differences between the control group and the performance on these examinations in the previous year ($.606 < p < 0.905$). Unfortunately, no means and standard deviations were provided for this study.

If academic ability, as assessed by university entry scores, plays a role in a student's decision to participate voluntarily in SI, one would want to know how to encourage students who enter with lower scores to participate in SI (Stansbury, 2001), and would SI be effective for this group. Using data from a pilot project in which students who, based on entry scores and other admissions information, were deemed to be at risk of failing a compulsory chemistry course, Stansbury (2001) concluded that those students felt they lacked the knowledge and self-efficacy to participate in SI. Hence, in a subsequent pilot, he introduced a brief intervention in which students received a pre-SI induction into the course. Of the 32 students who participated, 16 were identified as being at risk of failure. The results suggested that the at-risk students who attended the pre-SI sessions did attend SI more frequently and gained better final grades. The small numbers of participants involved in this intervention and who filled in the pre- and post-self-efficacy assessments did not allow for statistical tests.

Effectiveness for Traditionally Underrepresented Student Populations

SI was explicitly designed *not* to target students at risk but to target courses perceived to be difficult (Blanc et al., 1983; Martin & Arendale, 1993). Martin and Arendale (1993) used the analogy that SI is like a public health initiative, in that it targets all students enrolled in a course, differentiating it from a medical approach of diagnosis and remediation. However, the initial impetus for the introduction of SI was changes in the student body following the successes of the civil rights movement, and interest in the effectiveness of SI for particular bodies of students has been a consistent feature of the research. In many cases, researchers are interested in students from underrepresented population groups, such as first-in-family students, ethnic minority, and indigenous students. The question is whether the design features of SI are useful in enhancing the academic performance of students who may not have done very well in more traditional teaching/learning environments.

More traditional pedagogical approaches, which are often premised on students at university level being able to absorb, process, and make sense of large amounts of information through transmission-style delivery models, may be exactly the reason why certain groups have underperformed (van der Meer, Scott, & Neha, 2010). In this section, we will consider studies that have sought to assess the benefits of SI for underrepresented minorities (URMs) as well as students enrolled in developmental education studies.

Peterfreund et al. (2008), in a study encompassing 11 SI courses over a 13-year period, found that URM students were more likely to enroll in SI than non-URM students and that the achievement of URM students who participated in SI groups, as expressed in course grades, exceeded the achievement of both non-SI URM students as well as the non-SI students who were not identified as belonging to an URM group. The authors also pointed out that the differences did not always reach significance because of the small numbers involved.

In a related study (Rath et al., 2007), partially drawing on the same data, the authors reported on URM students attending an introductory biology course spanning the period between 1994–1998 (no SI provided) and 1999–2005 (SI provided). Both the mathematics and verbal SAT scores were lower for the SI participants than non-SI students (though only the verbal scores were significant). Apart from reporting that those students who did participate in SI earned considerably higher final grades (though no significance values are provided), they also found that the percentage of URM-SI participants who ultimately graduated was higher (73%) than for URM-non-SI participants who did not (50%). For non-URM students, no clear differences were found between SI participants (65%) and non-SI participants (62%).

The authors acknowledged that there were some data-related issues, as they drew mainly on institutionally available data. They also recognized that other interventions may have contributed to higher graduation rates for URMs. Rath et al. (2007) suggested a number of possible reasons why the results were so positive. These included the larger number of URM students at this particular university (resulting in lesser isolation), specific efforts made to encourage URM students to participate in SI, students being informed about SI by friends/family, and the consistency and quality of the SI organization.

Differential Effects of SI on Male and Female Students

Various studies sought to establish whether gender played a role in success rates for SI participation. The best designed study we found in this category was that of Fayowksi and MacMillan (2008). To establish the possible effect of gender on final course grade difference, they performed an analysis of covariance. Although there was a difference in results for the mean grades of male and female students, there was no significant interaction effect by gender for SI.

Peterfreund et al. (2008) compared the course grade differences between male and female students across a range of courses. They concluded that where male students participated in SI, the differences between them and male students who did not participate were greater than the differences between female SI participants and nonparticipants. Peterfreund et al.'s results indicate that male participation in SI did not exceed a third of total participation in SI. Hodges et al. (2001) commented that male students were underrepresented in participation in SI in their study but did not provide any information on differential effectiveness by gender.

Effectiveness Beyond the Course in Which SI Was Implemented

Few studies that met the inclusion criteria explored the effect of SI on graduation rates. A notable exception was the study by Bowles et al. (2008). They argued that single-equation model approaches cannot adequately calculate the effect of SI. In this study, Bowles et al. used a two-equation model with proxy indicators for motivation and ability to calculate the effect of SI attendance during the first year at college/university on graduation. They estimated that SI attendance increased the probability of timely graduation by nearly 11%. They pointed out, however, that this result had to be considered in the context of the demographic profile of their particular sample, a Utah university with a high population of Mormon students, many of whom interrupt their studies for 2 years of church service.

Ogden et al. (2003) tracked students for a full academic year after their initial participation in SI. They found that, whereas there was no noticeable effect on the quarterly GPA for traditional SI students, there was an effect for the conditional SI group. The authors suggested that students in this group might have benefited from a booster in SI. They also noted that students in the conditional non-SI group improved over the course of the academic year, which they speculated might be because these students slowly developed learning skills over time and because the weaker students may have left. Although the numbers are small, of particular note was their finding that conditional SI students had the highest percentage of reenrolment at the institution of all four groups. They suggested that this particular group may have been more motivated to do well and persist.

Students' choice to continue in a discipline could also be a function of interest awakened during SI sessions, or because of an experience of academic success. Miller et al. (2004) suggested that introduction of SI in a first-year chemistry course may have had a positive flow-on effect on enrolment in higher level chemistry courses. The somewhat confusing table presented in the results makes it difficult to interpret the strength of this flow-on effect. Although they recognized that it was difficult to claim that it was definitively SI that caused this improvement, there had been no change in overall enrolment numbers in the first-year course. They did not provide information on whether the course had changed in any other way. Peterfreund et al. (2008) also found that students who had participated in SI for biology or chemistry tended to continue with further courses in that discipline.

Bowles and Jones (2003–2004b) cautioned against drawing simplistic conclusions from the relationship between SI attendance and retention. In their study of 3,645 students, those who participated in SI seemed to return at a higher rate (89.9%) than those who did not attend (82.7%). Rather than assuming that return rate is caused by SI participation, it is important to acknowledge that the return rate could be the result of an unmeasured variable, for example, motivation. As these outcomes may be jointly determined, Bowles and Jones (2003–2004b) advocated measuring retention using a bivariate probit model.

Impact on Academic Skills Development

Most SI programs would claim that one of the benefits for students participating in SI is that they develop a range of academic skills, such as academic reading and note-taking, for example. Few studies, however, specifically report on these outcomes. Longfellow et al. (2008) reported that SI-students seemed to have developed better writing skills. This conclusion, they say, was evidenced through the actual assessments and student responses to survey questions. Ogden et al. (2003) reported on the results of a survey they administered at the completion of an SI program in an American university. They did not provide numbers, but indicated that students noted in particular the benefits of being exposed to effective note-taking and exam preparation techniques. Students in a U.K. program (Smith et al., 2007) were asked to fill in a survey that sought to ascertain, among other questions, whether participants had gained particular academic skills. The results from the small number of surveys returned (10 out of 35 attendees) did not provide convincing evidence in support of this claim. A greater understanding of the course expectations seemed to be the biggest gain.

Court and Molesworth (2008), in a British study, found that 70% of 34 respondents thought that getting a good understanding of the course expectation, or uncertainty reduction was the most valued aspect of SI. In a New Zealand study (van der Meer & Scott, 2009), the authors included a number of questions related to study skills (with 5 being *strongly agree* on a 5-point scale). The questions related to time management and workload yielded a mean score of 3.11 (SD = 0.96) and 3.03 (SD = 1.02) for note-taking. The authors concluded that this was not a convincing result. A more general question on whether respondents thought SI had helped them develop study and learning strategies received higher ratings (M = 3.72, SD = 0.89). However, as the authors remarked, the respondents may also have interpreted this to mean study skills directly related to content, rather than more general transferable academic skills.

Ning and Downing (2010) used the Learning and Study Strategies Inventory by Weinstein and Palmer (2002) to examine the effect of SI on academic skills development in a Hong Kong university. The Learning and Study Strategies Inventory consists of 10 subscales. SI participants showed more gains than non-SI participants on two of the 10 subscales, information processing skills (d = 0.43 for SI participants vs. d = 0.13 for non SI participants) and motivation levels (d = 0.39 vs. d = 0.10). The 10 subscales are used to create three latent constructs, self-regulation, skills, which Ning and Downing collapsed into one construct, learning competence. They showed that learning competence had a mediating effect on the achievement of both the SI participants and non-SI participants. They concluded that academic achievement is not just a function of learning competence, but that learning competence can be developed through SI.

Effect on General Satisfaction or Well-Being

It could be argued that any intervention that results in students doing better in their course will result in enhanced well-being. Some studies explicitly refer to this impact of SI. Bronstein (2008), who interviewed students and conducted a focus group, reported that SI participation in a particularly difficult and compulsory course, helped reduce students' anxiety. Reasons advanced for this outcome included being part of a supportive environment, and being able to discuss difficult material with other students.

Although not always explicitly stated, being able to enroll in a voluntary academically focused program in addition to faculty-organized teaching and learning sessions seemed to give students a sense that they had something extra: extra support and extra opportunities to engage in academically focused time on task. This extra dimension was often expressed in the context of being able to ask questions (Longfellow et al., 2008; Mahdi, 2006; van der Meer & Scott, 2009). Students involved in focus groups conducted by Dobbie and Joyce (2008), for example, reported on this extra benefit. These authors indicated that for the majority of the students in their focus group, the extra support provided by the SI session was the main reason why they found the sessions helpful. It is also possible that voluntary participation supported students' development of a sense of control over their learning, which is a small but important contributor to success in higher education (Richardson et al., 2012).

Sometimes this extra was experienced as a compensatory opportunity for something they did not get in the lectures or tutorials, either because of time, because of logistical constraints in class contexts (large numbers of students), or because of perceptions that the environment was not relaxed or safe enough to ask questions. Longfellow et al. (2008) reported that students felt intimidated and were more reluctant to ask questions in lectures. In the study by van der Meer and Scott (2009), 15% of the 286 respondents (out of 345) who filled in the open-text question identified being able to get help as a defining characteristic of SI, often referring to the ability to ask questions.

Conversely, however, the idea that SI is a support or help service sometimes mitigated against students enrolling in SI, arguably often for students who would most benefit from attending SI sessions. As Martin and Arendale (1993) stated, "Whether through denial, pride, or ignorance, students who need help the most are least likely to request it" (p. 42). This failure to enroll was one of the reasons why Hodges et al. (2001) experimented with compulsory SI attendance. They referred to studies that suggest a curvilinear pattern in help-seeking behavior, with the least able being less likely to participate in SI.

Enhanced Social Relationships

Participation in SI or similar peer-learning programs can enhance students' opportunity to meet other students and develop new friendships. Although not every SI program explicitly states these social benefits in their design intentions, students' evaluation of SI typically seemed to yield comments that support this benefit. Dobbie and Joyce (2008) conducted a number of focus groups with students who attended peer-learning programs. The results from their small qualitative project suggested that students appreciated this aspect of attending SI sessions. The authors remarked that students from abroad (non-Australian students) particularly valued the opportunity to make new friendships. They also emphasized the important role that the social aspect of SI can play in students' integration into the university.

In other studies, writers reported more general student remarks about the benefits of student to student contact in small groups (Mahdi, 2006). Hafer (2001) suggested that introduction of SI in her English composition class has enhanced peer responses; however, this finding is anecdotal. Similarly, Court and Molesworth (2008) reported that 11 of 33 survey respondents identified social gains as a benefit of participation in SI, whereas eight mentioned social gains as the best aspect of SI.

Effectiveness in Engagement

Few studies explicitly sought to establish whether SI provided students with a learning opportunity that was considered more engaged than traditional teaching/learning opportunities provided by faculty. Although Fayowksi and MacMillan (2008) did not set out to study this aspect of SI, they reported that for the three semesters during which SI was offered, students chose not to attend tutorials, although the faculty-organized tutorials were run by excellent tutors. This was also the case before introduction of SI, which was the reason, they added, for considering SI in the first place. Although the authors do provide the figures for SI participants who attended five or more sessions (269) versus those who attended no SI sessions (990), they do not provide comparable figures for tutorial attendance before or after introduction of SI.

Summary

In considering the range of approaches in the articles that met the inclusion criteria, it was found that a considerable number of studies did not provide all the details that would have allowed for a comprehensive assessment of their findings. For example, studies omitted definitions of what constituted SI attendance, number of students involved, *p* values, mean grades, and standard deviations. Absence of the latter two bits of information meant that effect sizes could not be calculated. The variability of methodologies, approaches, measures, and variable characteristics (e.g., grade scales) also prevented a more rigorous assessment, for example, a meta-analysis, and comparison of the claims made in the studies. In seven of the studies included in the review, it was possible to calculate the effect size of SI on final course grade as an indication of effectiveness of the treatment group compared with the group who had not participated in SI. These effect sizes ranged from $d = 0.29$ to $d = 0.60$, with an unweighted average effect size being $d = 0.48$. Five of these effects were above Hattie's 0.4 cutoff for interventions worth implementing (Hattie, 2009). The variability of studies, however, especially with regards to *controls for self-selection* needs to be kept in mind in considering this result.

Many studies used quasi-experimental study designs, and some authors suggest that studying SI using experimental designs will enhance the results in this area. Cheng and Walters (2009), although advocating for such an approach, sounded a note of caution because of the sheer number of variables that would need to be controlled. They also expressed some concerns about the ethical issues in denying some students the SI intervention. We found one study (Parkinson, 2009) that used random assignment of students to experimental and control groups.

A persistent question raised was why some students who might benefit do not participate in SI. Bronstein (2008), for example, wondered how we can help first- and second-year students appreciate the importance of availing themselves of support provided. Stansbury (2001) set up an intervention to address this issue and recommended that this intervention be repeated with a larger number of participants to confirm the tentative results in the small-scale study. Hodges et al. (2001) sought to respond to this concern by experimenting with making SI compulsory.

Related to the question of nonparticipation is the question of who participates in SI. In many studies, the authors recognized the problematic nature of one of the defining characteristics of the SI model, the voluntary nature of participation. This concern makes it difficult to assess whether achievement gains result from the SI treatment or from self-selection into the program, with the dominant assumption being that more motivated or academically able students may be more likely to participate. Bowles and Jones (2003–2004a), however, claimed that the traditional approach to account for self-selection using a single-equation ordinary least squares model (using preuniversity entry scores as a dependent variable) underestimates the effect of SI. In their study, they used a simultaneous equation (limited dependent variable model of SI effectiveness) and established a larger effect for SI after controlling for academic ability. Academically less able students in their study participated in larger numbers in SI. They did recognize, however, that their findings may reflect their particular cohort of students in a Utah-based university.

The included studies largely focused on achievement results as measured by final course grades. Although one study did measure the possible effects on study skills and dispositions such as motivation (Ning & Downing, 2010), there was a noticeable lack of quantitative studies addressing these outcomes. Equally absent were many studies that addressed the differential effect of SI on ethnic or under-represented minorities.

Ten studies employed some kind of qualitative method, either exclusively or as a complement to quantitative approaches. Data collection included surveys, interviews, and focus groups, and thematic analysis approaches were the most common means of interpretation. Although the methods of qualitative data collection were employed in about a third of our included articles, we did not find examples of sustained, theoretically informed, rigorous adherence to an overarching qualitative methodology (e.g., phenomenography, grounded theory, or ethnography).

Publication bias may have influenced the findings of our review. Studies with significant, positive, large effects tend to be more attractive to journal editors than studies finding no significant effect for an educational intervention (Torgerson, 2006). Although formal methods exist for detecting publication bias (such as the funnel plot or fail-safe n), they rely on formulaically comparing effect sizes. This process could not be undertaken due to the substantial diversity of methodologies employed and definitions of the SI intervention, as well as the inability to calculate effect sizes based on the limited information provided by most studies. We do however point out that the typical study showed some sort of positive result and was written by authors who were employed in a position where sharing this sort of result is advantageous and sharing a negative or insignificant result could be career limiting. It would be a courageous SI supervisor that published a result showing that SI did not help their students.

Conclusion

This review has been useful in surveying effectiveness studies related to SI for the 2001 to 2010 period for a number of reasons. First, it provides both practitioners and researchers in the field of peer learning with an updated inventory of literature on effectiveness related to this particular form of peer learning. The review may be particularly useful for those considering introducing SI in their institutions or responding to those in their institutions who question the empirical evidence base for introducing or continuing to support SI. Although, at best, we can say that in many instances, SI seemed to have been effective (keeping in mind possible publication bias), it does provide some indications that SI *worked* on some level for some groups of students. Although there is a marked lack of an evidence base in demonstrating effectiveness in the field of learning support interventions, SI has benefited from at least a degree of empirical research over more than 40 years in many different countries.

Second, the review has been useful in highlighting some possible improvements in conducting research into this model of peer learning, with four recommendations made to guide future research in this area. The first recommendation is that any future study provides the bare minimum of data that would enable other researchers to verify claims of effectiveness. This bare minimum would include (a) numbers of students involved in any group for which data are reported on, (b) SI attendance requirements for membership in a particular group, (c) the mean course grades, and (d)the range of course grades (minimum, maximum) as well as standard deviations. Furthermore, where comparisons between different achievement data are provided, significance levels (*p* values) should be included. The second recommendation is that researchers explore ways to include control for self-selection. This, however, is no easy task and may differ depending on the local context, for example, if a measure of preentry achievement is included as proxy for academic ability.

Third, this review has suggested that the field of SI research may benefit from cross-institutional and transnational research collaboration projects to start to identify particular ways in which SI is implemented that seem to produce more convincing results than alternative ways. Collaborations will also benefit institutions that are considering implementing this form of learning support intervention. This ability for collaborative and comparative research also alludes to the need for researchers to be clear about the institutional and discipline context in the reporting of any study.

Fourth, this review has clearly highlighted that certain aspects of SI effectiveness are underresearched. Effects of SI other than on academic achievement as expressed in course grades or course completion would add to the values institutions may attach to this model. Some studies do address social and transferable benefits. There is considerable room, however, for this focus of research to be broadened and made more robust. A more intentional focus on achievement of URMs too is desirable. In many countries, governments are concerned with equity goals for higher education. SI may be a very suitable candidate for this agenda to be advanced. Apart from the origins of SI that clearly place this program in a group of interventions that have proven successful outcomes for students in minority populations, the design features also may suit particular population groups with a more collectivistic and relational orientation rather than a more individualistic orientation.

In addition, with many countries experiencing the opening up of tertiary education to a widening student population, analysis of the impact of SI on other diverse student groups, such as low socioeconomic status, first in family, and international students, may be of interest. As institutions worldwide become increasingly interested in tracking their at-risk populations and analyzing the impact of different learning interventions (e.g., for quality or financial considerations), having rigorous research on the effectiveness of SI becomes increasingly important. Apart from being able to provide support for these targeted interventions, evidence of SI effectiveness can also be a considered in the context of evaluation of a wider range of programs aimed at enhancing the transition into higher education and the first-year experience of all students. SI is one piece of the transition puzzle that deserves attention, given its longevity and breadth of application across the globe.

Returning to the question that motivated this study—"What is the effectiveness of Supplemental Instruction for the attending students?"—we have found a diverse range of effects. Each of the specific claims validated by the U.S. Department of Education is supported by published research from 2001 to 2010, and none is contradicted by research from this period. Importantly, however, none is supported by a gold standard study involving random assignment to groups and sufficient detail about methodology, participants, and the SI intervention in practice. Also missing from the reviewed articles was a rigorous qualitative study employing a clear methodology that is well grounded in learning theory. Both types of studies will be necessary to fully address the research question in the future. It is possible that such studies have been published before or after our review period, but the authors of this study are not aware of them.

Note

Special thanks to Sally Rogan, University of Wollongong, Australia, for her support of this project, and to research assistants, Brittany Reeves and Andrew Milne.

References

Arendale, D. (1994). Understanding the supplemental instruction model. In D. Martin & D. Arendale (Eds.), *Supplemental instruction: Increasing achievement and retention* (*Vol. 60*, pp. 11–21). San Francisco, CA: Jossey-Bass.

Arendale, D. (1997). Supplemental instruction (SI): Review of research concerning the effectiveness of SI from the University of Missouri-Kansas City and other institutions from across the United States. In S. Mioduski & G. Enright (Eds.), *Proceedings of the 17th and 18th Annual Institutes for Learning Assistance Professionals: 1996 and 1997*. Tucson: University Learning Center, University of Arizona.

Arendale, D. (2002). History of supplemental instruction (SI): Mainstreaming of developmental education. In D. B. Lundell & J. Higbee (Eds.), *Histories of developmental education* (pp. 15-27). Minneapolis: Center for Research on Developmental Education and Urban Literacy, General College, University of Minnesota.

Ashwin, P. (2003). Peer support: Relations between the context, process and outcomes for the students who are supported. *Instructional Science*, *31*, 159–173. doi:10.1023/A:1023227532029

Baker, R., & Siryk, B. (1984). Measuring academic motivation of matriculating college freshmen. *Journal of College Student Personnel*, *25*, 459–464.

Bearman, M., & Dawson, P. (2013). Qualitative synthesis and systematic review in health professions education. *Medical Education*, *47*, 252–260. doi:10.1111/medu.12092

Blanc, R., DeBuhr, L., & Martin, D. (1983). Breaking the attrition cycle: The effects of supplemental instruction on undergraduate performance and attrition. *Journal of Higher Education*, *54*, 80–90. doi:10.2307/1981646

Bowles, T. J., & Jones, J. (2003–2004a). An analysis of the effectiveness of supplemental instruction: The problem of selection bias and limited dependent variables. *Journal of College Student Retention: Research, Theory and Practice*, *5*, 235–243. doi:10.2190/486t-mhvc-cg0c-rm3b

Bowles, T. J., & Jones, J. (2003–2004b). The effect of supplemental instruction on retention: A bivariate probit model. *Journal of College Student Retention: Research, Theory and Practice*, *5*, 431–437. doi:10.2190/jg2q-vnxr-k6gq-n810

Bowles, T. J., McCoy, A. C., & Bates, S. C. (2008). The effect of supplemental instruction on timely graduation. *College Student Journal*, *42*, 853–859.

Bradley, D., Noonan, P., Nugent, H., & Scales, B. (2008). *Review of Australian higher education: Final report*. Canberra, Australia: Department of Education, Employment and Workplace Relations.

Bronstein, S. B. (2008). Supplemental instruction: Supporting persistence in barrier courses. *Learning Assistance Review*, *13*, 31–45.

Cheng, D., & Walters, M. (2009). Peer-assisted learning in mathematics: An observational study of student success. *Journal of Peer Learning*, *2*, 23–39.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Congos, D., & Mack, A. (2005). Supplemental instruction's impact in two freshman chemistry classes: Research, modes of operation, and anecdotes. *Research & Teaching in Developmental Education*, *21*(2), 43–64.

Court, S., & Molesworth, M. (2008). Course-specific learning in peer assisted learning schemes: A case study of creative media production courses. *Research in Post-Compulsory Education*, *13*, 123–134. doi:10.1080/13596740801903729

Crisp, G., & Cruz, I. (2009). Mentoring college students: A critical review of the literature between 1990 and 2007. *Research in Higher Education*, *50*, 525–545. doi:10.1007/s11162-009-9130-2

Dancer, D., Morrison, K., & Smith, M. (2007). Measuring the impact of a peer assisted learning program on students' academic performance in econometrics. In M. Davies, J. Hirschberg, C. Johnston, & J. Lye (Eds.), *The quantitative analysis of teaching and learning in higher education: Forum proceedings* (pp. 19–42). Melbourne, Victoria, Australia: University of Melbourne.

Dixon-Woods, M., Agarwal, S., Jones, D., Young, B., & Sutton, A. (2005). Synthesising qualitative and quantitative evidence: A review of possible methods. *Journal of Health Services Research & Policy*, *10*(1), 45B–53B.

Dobbie, M., & Joyce, S. (2008). Peer-assisted learning in accounting: A qualitative assessment. *Asian Social Science*, *4*(3), 18–25.

Fayowski, V., & MacMillan, P. D. (2008). An evaluation of the supplemental instruction programme in a first year calculus course. *International Journal of Mathematical Education in Science and Technology*, *39*, 843–855. doi:10.1080/00207390802054433

Gattis, K. W. (2002). Responding to self-selection bias in assessments of academic support programs: A motivational control study of supplemental instruction. *Learning Assistance Review*, *7*(2), 26–36.

Gehanno, J.-F., Rollin, L., & Darmoni, S. (2013). Is the coverage of Google Scholar enough to be used alone for systematic reviews. *BMC Medical Informatics and Decision Making*, *13*, 7. doi:10.1186/1472-6947-13-7

Giustini, D., & Boulos, M. N. K. (2013). Google scholar is not enough to be used alone for systematic reviews. *Online Journal of Public Health Informatics*, *5*(2). doi:10.5210/ojphi.v5i2.4623

Hafer, G. R. (2001). Ideas in practice: Supplemental instruction in freshman composition. *Journal of Developmental Education*, *24*(3), 30–37.

Hattie, J. (2009). *Visible learning: A synthesis of 800+ meta-analyses on achievement*. Oxford, England: Routledge.

Hensen, K. A., & Shelley, M. C. (2003). The impact of supplemental instruction: Results from a large, public, Midwestern university. *Journal of College Student Development*, *44*, 250–259. doi:10.1353/csd.2003.0015

Hodges, R., Dochen, C., & Joy, D. (2001). Increasing students' success: When supplemental instruction becomes mandatory. *Journal of College Reading and Learning*, *31*, 143–156.

Hodges, R., & White, W. G. (2001). Encouraging high-risk student participation in tutoring and supplemental instruction. *Journal of Developmental Education*, *24*(3), 2–43.

International Center for Supplemental Instruction. (2003). *Supplemental instruction: National data summary, 1998–2003*. Retrieved from http://www.umkc.edu/asm/si/si-docs/NationalSupplementalInstructionReport98-03.pdf

Kochenour, E., Jolley, D., Kaup, J., Patrick, D., Roach, K., & Wenzler, L. (1997). Supplemental instruction: An effective component of student affairs programming. *Journal of College Student Development*, *38*, 577–586.

Longfellow, E., May, S., Burke, L., & Marks-Maran, D. (2008). "They had a way of helping that actually helped": A case study of a peer-assisted learning scheme. *Teaching in Higher Education*, *13*, 93–105. doi:10.1080/13562510701794118

Mahdi, A. E. (2006). Introducing peer-supported learning approach to tutoring in engineering and technology courses. *International Journal of Electrical Engineering Education*, *43*, 277–287. doi:10.7227/ijeee.43.4.1

Martin, D. (2009). Foreword. *Australasian Journal of Peer Learning*, *1*(1), 3–5.

Martin, D., & Arendale, D. (1993). *Supplemental instruction: Improving first-year student success in high-risk courses* (2nd ed.). Columbia: National Resource Center for the First Year Experience and Students in Transition, University of South Carolina.

Martin, D., Arendale, D., & Blanc, R. (1997). *Mainstreaming of developmental education: Supplemental instruction and video-based supplemental instruction*. Kansas City, Missouri: National Center for Supplemental Instruction, University of Missouri-Kansas City. Retrieved from http://a.web.umkc.edu/arendaled/mainstreamDE97.pdf

McCarthy, A., Smuts, B., & Cosser, M. (1997). Assessing the effectiveness of supplemental instruction: A critique and a case study. *Studies in Higher Education*, *22*, 221–231. doi:10.1080/03075079712331381054

Miller, V., Oldfield, E., & Bulmer, M. (2004). Peer assisted study sessions (PASS) in first year chemistry and statistics courses: Insights and evaluations. In I. Johnston (Ed.), *2004 National UniServe Conference* (pp. 30–35). *Sydney, New South Wales, Australia: UniServe Science*.

Moore, R., & LeDee, O. (2006). Supplemental instruction and the performance of developmental education students in an introductory biology course. *Journal of College Reading and Learning*, *36*(2), 9–20.

Ning, H. K., & Downing, K. (2010). The impact of supplemental instruction on learning competence and academic performance. *Studies in Higher Education*, *35*, 921–939. doi:10.1080/03075070903390786

Ogden, P., Thompson, D., Russell, A., & Simons, C. (2003). Supplemental instruction: Short -and long-term impact. *Journal of Developmental Education*, *26*(3), 2–8.

Øvretveit, J., & Gustafson, D. (2002). Evaluation of quality improvement programmes. *Quality and Safety in Health Care*, *11*, 270–275. doi:10.1136/qhc.11.3.270

Parkinson, M. (2009). The effect of peer assisted learning support (PALS) on performance in mathematics and chemistry. *Innovations in Education and Teaching International*, *46*, 381–392. doi:10.1080/14703290903301784

Pawson, R. (2006). *Evidence-based policy: A realist perspective*. London, England: Sage.

Peterfreund, A. R., Rath, K. A., Xenos, S. P., & Bayliss, F. (2008). The impact of supplemental instruction on students in STEM courses: Results from San Francisco

State University. *Journal of College Student Retention: Research, Theory and Practice*, *9*, 487–503. doi:10.2190/CS.9.4.e
Phelps, J. M., & Evans, R. (2006). Supplemental instruction in developmental mathematics. *Community College Enterprise*, *12*(1), 21–37.
Rath, K. A., Peterfreund, A. R., Xenos, S. P., Bayliss, F., & Carnal, N. (2007). Supplemental instruction in Introductory Biology I: Enhancing the performance and retention of underrepresented minority students. *CBE-Life Sciences Education*, *6*, 203–216. doi:10.1187/cbe.06-10-0198
Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, *138*, 353–387. doi:10.1037/a0026838
SI Staff from UMKC. (2005). *Supplemental instruction supervisor manual*. Kansas: University of Missouri-Kansas City.
SI Staff from UMKC. (2010). *Supplemental instruction/Video supplemental instruction annotated bibliography*. Retrieved from http://www.umkc.edu/asm/si/si-docs/sibib.htm
Smith, J., May, S., & Burke, L. (2007). Peer assisted learning: A case study into the value to student mentors and mentees. *Practice and Evidence of Scholarship of Teaching and Learning in Higher Education*, *2*(2), 80–109.
Stansbury, S. (2001). Accelerated learning groups enhance supplemental instruction for at-risk students. *Journal of Developmental Education*, *24*(3), 20–40.
Torgerson, C. J. (2006). Publication bias: The Achilles Heel of systematic reviews? *British Journal of Educational Studies*, *54*, 89–102. doi:10.1111/j.1467-8527.2006.00332.x
U.S. Department of Education. (1995). Supplemental instruction (SI): Improving student performance and reducing attrition. In G. Lang (Ed.), *Educational programs that work* (21st ed., p. 14.4). Longmont, CO: Sopris West.
van der Meer, J., & Scott, C. (2009). Students' experiences and perceptions of peer assisted study sessions: Towards ongoing improvement. *Australasian Journal of Peer Learning*, *2*(1), 3–22.
van der Meer, J., Scott, S., & Neha, T. (2010). Retention of first-year Māori students at university. *MAI Review*, *2010*(2). Retrieved from http://www.review.mai.ac.nz/index.php/MR/article/view/309
Weinstein, C. E., & Palmer, D. R. (2002). *LASSI user's manual* (2nd ed.). Clearwater, FL: H&H Publishing.
Wright, G. L., Wright, R. R., & Lamb, C. E. (2002). Developmental mathematics education and supplemental instruction: Pondering the potential. *Journal of Developmental Education*, *26*(1), 30–35.

Authors

PHILLIP DAWSON holds a first-class honors degree in computer science and a PhD in education. He is a lecturer in learning and teaching in the Office of the Pro Vice-Chancellor (Learning and Teaching) at Monash University, PO Box 197, Caulfield East, Victoria 3145, Australia; e-mail: *phillip.dawson@monash.edu*. His research focuses on the ways that university teachers use a variety of evidence to inform their teaching.

JACQUES VAN DER MEER holds a master's degree in education (endorsed in counseling) with first-class honors from the University of Auckland and a PhD from the University of Otago. He is the associate dean (academic) at the University of Otago

College of Education, Otago, New Zealand; e-mail: *jacques.vandermeer@otago.ac.nz*. His research areas relate to student transition and induction into higher education, student engagement, and peer learning. He also has an interest in student retention and achievement, especially of underrepresented minorities.

JANE SKALICKY holds a first-class honors degree and a PhD in education. She is head of student learning and academic development at the University of Tasmania, Hobart, Tasmania, Australia; e-mail: *jane.skalicky@utas.edu.au*. Her research focus is in learning and teaching in higher education, specifically peer learning, as well as numeracy education within both school and tertiary sectors.

KYM COWLEY has a PhD in management and first class honors degree in marketing. She is director peer learning and student progression, Faculty Business and Law, University of Newcastle, New Castle, New South Wales, Australia; e-mail: *kym.cowley@newcastle.edu.au*. Her research focus is peer learning and intensive learning methods in higher education.