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

Phillip Dawson

Monash University

Jacques van der Meer

University of Otago

Jane Skalicky

University of Tasmania

Kym Cowley

University of Newcastle

Author note

Phillip Dawson, Office of the Pro Vice-Chancellor (Learning and Teaching), Monash University, Australia

Jacques van der Meer, College of Education, University of Otago, New Zealand

Jane Skalicky, Student Learning & Academic Development, University of Tasmania, Australia

Kym Cowley, Newcastle Business School, University of Newcastle, Australia

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

Correspondence concerning this article should be sent to Phillip Dawson, Office of the Pro Vice-Chancellor (Learning and Teaching), Monash University, PO Box 197, Caulfield East, Victoria, 3145, Australia. E-mail: phillip.dawson@monash.edu

Abstract

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 paper 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 7 included studies, however, an effect size of SI participation on final grades was calculated, ranging from $d = 0.29$ to $d = 0.6$. The findings of the review are consistent with claims validated by the US 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

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

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 is in the tens of million of dollars. Staff from more than 1500 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 2001-2010 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), but 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 STEM subjects, although the 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 ‘re-teach’ 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 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: United States of America (USA); Canada; United Kingdom (UK); Europe; South Africa; and Australasia. Whereas the undergraduate student mentoring literature finds 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 email list.

Supplemental Instruction 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 to the grades of non-participants, 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 US Department of Education in 1992 (Martin & Arendale, 1993). Their 'validation' of SI has been cited hundreds of times in the scholarly literature, and contains these 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 is still true when analysis controls 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 (re-enroll and graduate) at higher rates than students who do not participate in SI.

(Martin & Arendale, 1993, p. 26 citing US Department of Education, 1992)

Claims 1 and 2 are supported by data analyzed by the UMKC team from 49 US institutions, representing 1,447 individual courses and an undisclosed number of students (Martin & Arendale, 1993). Three studies (n = 1,689; 349; 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, criticize earlier attempts to assess the effectiveness of SI. Treatment of SI attendance as a binary variable (students either attended or did not attend) is described as simplistic, with McCarthy, et al. preferring a discrete variable for the number of sessions attended. They also question the usefulness of pre-entry test scores as a proxy for self-selection, as they claim such scores may not necessarily be correlated with success in tertiary study. The results of the McCarthy, et al. (1997) 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 (for example, Bowles & Jones, 2003-2004b; Hensen & Shelley, 2003; Hodges, Dochen, & Joy, 2001) do not cite the work of McCarthy, et al., and appear to be unaware of the issues they raise. Bowles, McCoy and Bates (2008) for example, who do cite the work of McCarthy, et al., (albeit incorrectly referenced), still view motivation as a function of prior academic achievement, and still treat 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) argues 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" approaches to their studies. The qualitative component of his study found 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" (p. 159).

Learning and grades are hopefully related, but it is naive to assume they are the same thing; 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 makes the methodological argument that diverse mixed-methods evaluations of peer learning programs are necessary to understand their true effectiveness.

Alongside Ashwin's critique of measures of effectiveness, and McCarthy et al.'s critique of measures for self-selection, is Kochenour, et al.'s (1997) critical review of the literature, which finds "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" (p. 578). 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 (for example, 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 non-systematic meta-analyses of unpublished results (for example, Arendale, 1997); they also were conducted by organizations who 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 non-essential 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

Supplemental Instruction 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 UK it is sometimes referred to as Peer Assisted Learning (PAL). 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*. We acknowledge 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

For the initial search, research assistants were instructed to exclude articles if they were obviously not about Supplemental Instruction 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 papers, with each paper 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. While 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 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 to be 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.

While acknowledging that there is a diversity of opinion within the practitioner community about what precisely distinguishes SI from other programs, determining if an article actually discussed SI was based on these criteria alone.

Data Extraction

Each included article was read by all four researchers, and structured data were extracted from each paper. In addition to basic details about each study, data extraction focused upon:

- The institutional and course context of each study
- Characteristics of the participants
- The research approach taken

- Any deviations from or augmentations to SI
- Claims about the effectiveness of SI, and 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 dataset 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.

Piloting

Inclusion/exclusion and data extraction processes were piloted and refined iteratively: firstly 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.

Disagreement and Conflicts of Interest

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

The initial search strategy yielded 596 results from databases and 819 results from Google Scholar. Research assistants included 98 of these articles based upon 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 included based upon 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 in Table 1 below.

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 all made reference to the literature, usually key works by Martin, Arendale or Blanc (most often Blanc, DeBuhr, & Martin, 1983; Martin & Arendale, 1993; Martin, Arendale, & Blanc, 1997). The use of the terms 'Supplemental Instruction' was not necessary for articles to be included, however the term appeared in every included paper 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 a program is SI because it is called SI would be 'label naivety' (Øvretveit & Gustafson, 2002, in Pawson, 2006) 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 & 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; as an example of a study with an SI session description, Rath, et al. (2007) describe:

> 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. (p. 207)

When claims of what happened in the SI sessions were given, like Rath's, they were usually not accompanied by evidence, and they 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 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 non-participants, 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: one 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); three sessions (Bowles & Jones, 2003-2004a, 2003-2004b; Bowles et al., 2008; Wright, Wright, & Lamb, 2002); five sessions (Fayowski & MacMillan, 2008); or twelve 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 (for example, Fayowski & MacMillan, 2008; Stansbury, 2001). Sometimes students were divided into groups of attendance, such as attending 0, 1-3, 4-6, 7-9, 10-12 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); given the diversity of definitions above we can only guess that one of those apply, most likely attending one or more sessions. In five other papers, 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). It is of concern that these particular papers 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 US 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—for example subject areas, and/or other areas of student benefit. With regards to the latter, there are a range of studies with additional effectiveness claims, for example research that sought to provide evidence of the effectiveness of SI in the development of academic skills and evidence for enhanced peer relationships.

In this section, an overview of effectiveness will be provided along these various categories of claims. The most common measures of effectiveness that were reported on are described first: final course grades and course completion rates. This is followed by studies that controlled for a range of factors, for example prior achievement (often as a proxy for ability), motivation, and college 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 analysis of the effectiveness of SI was based upon final course grades; 16 studies used this dependent variable. In many cases this was done 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 SI or did not opt into SI. Or formulated differently, as most SI programmes are voluntary, the studies used self-assigned treatment and control groups. Where authors sought to consider course grade differences in the context of factors such as motivation and/or ability, these studies will be discussed later, in the section on *controls for self-selection*.

Overall it is noted that most studies tested for significance between the means of the two groups (SI participants and non SI participants) and found that these differences were statistically significant. Many studies did not provide effect sizes, or standard deviations (that would have allowed effect sizes to be calculated). Studies emanating from Australia typically used marks out of 100 (e.g. Dancer et al., 2007; Miller et al., 2004), studies from the US typically used a 4-point scale. As an example of some of these typical studies, Dancer, et al. (2007) report on an Econometrics course with 628 students not attending SI and 262 attending. The difference between the mean marks, 58.9 versus 64.1 was significant at the $p<.001$ level and had an effect size of $d = 0.39$. Hensen and Shelley's study (2003) into four science and mathematics courses established significant difference between SI and non-SI participants at the $p<.05$ level, but did not provide exact significance levels. A study on the results of SI over a period of thirteen years in Science, Technology, Engineering and Mathematics (STEM) courses (Peterfreund et al.,

2008) reported significant differences ($p<.005$) for most of the courses. Whilst most SI programmes operate in first-year courses only, SI in this university was also included in some upper level courses. In the entry level courses, the authors noted that SI contributed to increased pass rates, and in higher level courses contributed to increased numbers of A's and B's. 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. This was confirmed for significance through analysis of variance ($F = 26.8$, $p<.0005$). They argued that using a 12-point scale enhanced the sensitivity of the analysis, however the creation of this 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, they say, for second semester scores to be lower than first semester scores. The scores of students who had attended SI sessions, however, 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, is 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 do not discuss these results. This makes 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, showed a graph with improvement of students who attended four sessions or more. However, as no numerical data was provided, it is difficult to assess the statistical significance of the results. An Australian study related to a first-year chemistry course (Miller et al., 2004) noted a difference between the mean results of students who attended no SI sessions (4.51), and students who attended more than five sessions (5.29). Although they make a claim that a Tukey post hoc test established a significant difference ($p<.05$) between the SI group who attended more than five sessions, and the group who attended 1-4 session and no sessions (but no difference between the students who attended 1-4 or no sessions), insufficient data is provided to verify this claim. This study is further hampered by absence of data on the number of students who attended SI (out of the 1131 enrolled students), or the exact grading scale used (it can be inferred that this could possibly be 1-7). Of interest in this study is that since the introduction of SI, the average GPA of the chemistry class improved from 4.34 in the two years prior to implementation, to 4.63, 4.61 and 4.91 since implementation. Whilst a positive impact, as the classes from year to year are comprised of different groups of students, some analysis in relation to self-selection would have strengthened these outcomes.

We found one study (Hodges & White, 2001) that compared SI to another form of learning support, which seem to have taken a more ‘conventional’ focus and was referred to as ‘tutoring’ (in the Northern American understanding of the term). 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. Whereas the final results for SI and non-SI participants, mean grades of 2.35 and 1.95 respectively, showed a significant difference, this was not the case for students who attended or did not attend tutorials, 2.23 and 2.10 respectively. While effect sizes were not published in this study, the researchers were able to calculate these from data provided, being $d = 0.51$ versus $d = 0.16$.

Although many of the studies reported final course mark comparisons, not all reported significance levels (e.g. Phelps & Evans, 2006; Wright et al., 2002) or student numbers for the different groups (e.g. Phelps & Evans, 2006). Whereas the majority of studies did not report effect sizes, many studies were also lacking standard deviations which would have enabled readers to calculate these.

Course completion. Whereas the main focus of some studies was on whether students did better in the course as expressed in final grades related to its completion, other studies focused on, or included, the pass and failure rates for courses.

Comparisons between the pass and failure groups, however, were carried out with varying levels of details. Whereas some studies merely provided proportions of pass (or grouped A, B and C results—ABC) and failures (D, F results and withdrawals—DFW), others also provided significance levels. In the study by Peterfreund, et al. (2008), pass levels for SI participants in most of the courses reached significance at the $p<.05$ level, and some at the $p<.005$ level. Congos and Mack (2005) provided only percentages of SI and non-SI ABC and

DFW groupings. Hensen and Shelley (2003) performed a Chi square test and reported that SI participants had significantly better pass rates at the $p<.05$ level, but did not provide exact significance levels.

Comparison before and after introduction of SI also figured in this category of effectiveness reporting. This could be seen as providing some degree of controlling for self-selection. One could, for example, argue that self-selection bias related to motivation or ability is likely to be limited if the successful completion rates of students *before* introduction of SI are not too dissimilar to successful completion rates of the group of students who chose not to enroll in SI *after* introduction of an SI programme. Congos and Mack (2005) reported on the pass and failure rates of two chemistry courses. They seem to make the case that the average DFWs (D and F grade and withdrawals) for SI participants were much lower than for non-SI students in one of the two courses (33% and 54% respectively). Their aggregated figures, however, may mask a more complicated picture. In the particular course on which they reported, the average rate of DFWs for students not enrolled in SI seemed to be on average higher than before introduction; this might possibly suggest self-selection of those who did participate in SI, that is: students who may have been less academically prepared for the course. Also, in one of the semesters, the failure rate of students enrolled in SI was higher than those who were not enrolled in 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 pre-treatment), but did so in a more systematic way. A Chi-square test suggested that the SI group did significantly better ($p<.0005$). The proportions of students succeeding and failing in the three groups, suggests, they say, that there was no significant difference between the non-SI group and the pre-treatment group (the year immediately before SI was introduced): 73% passing vs 27% failing for the SI group, 53% passing vs 47% failing for the non-SI group and 47% passing vs 53% failing for the pre-treatment group. They used this similarity as justification for combining the pre-SI and non-SI groups in their subsequent analyses.

This category of studies (*course completion*) also includes those that considered the probability of students passing a course 'because' of attendance in an SI programme. Cheng and Walters (2009) studied the chance of successful course completion of 534 students (out of a total 816) enrolled in two mathematics courses. They performed a logistic regression including 16 possible variables that might predict students' success. They found an odds ratio of 1.21 for each SI attendance. They caution, however, that over a third of the records had missing data and could therefore not be included.

Assessment tasks performance. Three studies sought to identify improvements in particular course work assessments other than just reporting final grades. Bronstein (2008), for example, claimed that assessments such as homework assignments and tests in a particular course also showed improvement because of SI. No data, however, was provided for this claim. Miller, et al. (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 practicals, multiple choice question exams and computer generated tests against students' SI attendance data. Differences between students not attending any SI sessions and students attending five or more sessions showed increased performance for the three assessment tasks: from 90.6 to 93.2 for assessment task one, from 50.2 to 61.0 for assessment task two, and from 74.3 to 83.9 for assessment task three. As mentioned above, however, not enough information was provided to assess these data fully on their merits. An Irish study reporting on the introduction of SI in a first-year Biotechnology course (Parkinson, 2009) provides achievement data for four successive course tasks. In comparing the results of the experimental (n = 24) and control group (n = 43) the results reveal a progressively significant difference between the first assessment task and 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 participate 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 programme based on motivation, after all SI 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 academically more 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 US 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 statistically more significant factor of the two variables that explained 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 upon, 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 ANCOVA results established that SI explained the difference in achievement and not the prior GPA. 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 an effect size of $d = 0.5$, which, can be considered as moderate (Cohen, 1988), and 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, whilst controlling for ability/motivation, Fayowski and MacMillan performed both a sequential logistic regression (significance at $p < 0.0005$) 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 participating in SI (with a minimum attendance of at least five sessions) was 2.7 greater than that for non-participants; one unit increase in prior GPA (on a twelve 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 by performing an ANOVA. 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 pre-entry characteristic by way of an ANCOVA suggested that SI participants did better ($p<.05$). Although they did not provide specific significance levels, Hensen and Shelley did provide details of standard deviations of the mean grades. This enabled us 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 ($p<.01$), but that there was no significant difference in the course marks. Moore and LeDee studied the results of four semesters with an average of 37 students attending SI, and 361 not. They considered SI successful, because 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). These behaviors they considered to be related to motivation. Moore and LeDee analyzed these academic behaviors for SI participants who earned an A or B grade and identified 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 pre-entry 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 lower 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 (UAI), which is a percentile rank. They found marginal differences of the UAI between SI and non SI participants (92. versus 91.8), and the correlation between UAI and SI participation was very small $r = 0.07$. Dancer, et al. conclude from this that brighter students did not necessarily attend 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 a Predicted Grade Point Average, or PGA, was calculated (Ogden et al., 2003), that included both SAT scores, high school marks and achievement prior to SI being offered. In this study by Ogden, et al. (2003), 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 PGA 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, but that higher grades were gained by students who attended more SI sessions. The 2606 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 programme 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). Their 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 lastly a group of students who were mandated to attend SI sessions (n = 108). This resulted in final mean grades of 2.49, 2.13 and 2.74, which they reported as significant. Post hoc analysis between non-SI and the other two groups revealed significance levels of $p = .0136$ for the voluntary group and $p = .0000$ for the mandatory group. 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. On first glance, then, 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, Dochen and Joy asked students to complete an academic motivation scale self-assessment (Baker & Siryk, 1984). This revealed a significant difference between voluntary SI participants (more motivated) and the mandatory SI group ($p = .0002$), but considerably less so between the non-SI and mandatory SI group ($p = .0443$). Comparison of SAT scores also revealed no significant difference, 871 for the voluntary group, 904 for the non-SI group and 895 for the mandatory group. For good measure they also looked at high school ranking: the rank score for the voluntary group was higher ($p = .03$) compared to the two other groups. These results might leave us with the conclusion that, although the mandatory group was less motivated and had lower pre-entry 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 session attended. They reported that the mandatory group had to attend all 13 sessions; this was not the case for the voluntary group. Few students in this latter group, they say, attended more than 10 sessions.

A further variation of controlling for self-selection was by 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 vs 2.24, and also better than the control group 2.70 (n = 11). We calculated an effect size of these as being *d* = 0.46 and d =0.14. The author subsequently adjusted (least squared means) the grades for pre-entry characteristics as measured by the Admissions Index (see above). This resulted in adjusted grades for SI participants of 2.96, non-participants 2.04 and control group 2.48. Effect sizes could not be calculated as no standard deviations were provided for this.

One study (Parkinson, 2009) created an experimental and control group from a first year intake of Biotechnology students. All students in this intake (n = 67) had volunteered for SI, but only 24, who were deemed to be representative of the composition in terms of pre-entry characteristics of the whole class, were assigned to the group who would receive the SI treatment. Differently from other SI programmes, SI sessions were connected to the whole programme of study rather than individual courses. Performance on four assessment tasks over the whole semester (see above) as well as three end of semester examinations showed significant differences between the experimental and control group (*p* values of 0.006, 0.003 and 0.001), and no significant differences between the control group and the performance on these examinations in the previous year (*p* values of 0.606, 0.75 and 0.905). Unfortunately no means and standard deviations were provided for this study.

If academic ability, as assessed by university entry scores, would play a role in a student's decision to participate voluntarily in SI, the question could be asked how students who enter with lower scores could be encouraged to participate in SI (Stansbury, 2001), and would SI be effective for this group? From data of 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 whereby students received a pre-SI induction into the course. Of the 32 students who participated in this pilot, 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 those who filled in the pre and post self-efficacy assessments, did not allow for statistical tests.

Effectiveness for traditionally under-represented student populations. Whereas SI was explicitly designed *not* to target students 'at risk' but perceived as difficult courses (Blanc et al., 1983; Martin & Arendale, 1993), it stands to reason that some academic and learning support staff will have considered whether a successful intervention like SI could be particularly useful for students who may not be as well prepared for university study. In many cases these students come from under-represented population groups such as first-in-family students, ethnic minority and indigenous students. The question is whether the design features of SI could be particularly 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, that 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 under-performed (van der Meer, Scott, & Neha, 2010). It is also noteworthy that SI was initiated in the 1970s to cater precisely for a group of students who were typically not familiar with university environments, African-American students, and who had started to enter universities as a result of the success of the civil rights movement. In this section, we will consider some of the few studies that have sought to identify the benefits of SI on under-represented minorities (URMs) as well as students enrolled in developmental education studies.

Peterfreund, et al. (2008) in a study encompassing data from eleven SI courses over a thirteen-year period, found that under-represented minority (URM) students were more likely to enroll in SI than non-URM students (except for two courses, this was at the $p<.005$ level) and that there was a consistent pattern of achievement of URM students who participated in SI groups. The achievement of these 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 under-represented minority group. The authors point 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 upon the same data, the authors report on the results of an introductory biology course, spanning the period between 1994-1998 (no SI provided) and 1999-2005 (SI provided), for URM students. For the SI participants, both the mathematics and verbal SAT scores were lower than those of the 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 those 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 for URM may have contributed to higher graduation rates. Rath, et al. 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 lastly, the consistency and quality of the SI organization.

Differential effect on male/female students. Various studies sought to establish whether gender played a role in success rates for SI participation, course grades and course completion. 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 ANCOVA. Although there was a difference in results for the mean grades of male and female students, there was no significant interaction effect for SI and gender. 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 non-participants. Their figures reveal that male participation in SI across courses evaluated, did not exceed a third of total participation in SI. Hodges, Dochen and Joy (2001) did comment on male students participation in SI as being under-represented in their study, but did not provide effectiveness differences.

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, McCoy and Bates (2008). Similar to other studies performed by Bowles and Jones (2003-2004a, 2003-2004b), they argued that single equation model approaches cannot adequately calculate the effect of SI. In the 2008 article, Bowles, McCoy and Bates used a two-equation model, using proxy indicators for motivation and ability, to calculate the effect of SI attendance during the first year at college/university on graduation. They calculated that this increased the probability of timely graduation by nearly 11%. They point 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 two 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 (for the differentiation of these two groups see the section above on motivational/ability differences). This SI effect, if this was indeed an effect, 'wore off' after that. 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 re-enrolment at the institution of all four groups. They suggest 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 chemistry first-year 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 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; for biology this was at the $p<.005$ level.

Bowles and Jones (2003-2004b) caution drawing simplistic conclusions from the relationship between SI attendance and retention. In their study of 3645 students, those who participated in SI seemed to return in greater proportion than those students who did not attend (89.9 versus 82.7%). Rather than assuming that one is caused by the other, this could be the result of an unmeasured variable, for example motivation. As these may be jointly determined variables therefore, they advocate to measure retention using a ‘bivariate probit model’.

Impact on academic skills development. Most SI programmes would claim that one of the benefits for students participating in SI is that students develop a range of academic skills, such as academic reading, note-taking etc. Few studies, however, specifically reported on this. Longfellow, et al. (2008) reported that students seemed to have developed better writing skills. This, 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 programme 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 UK programme (Smith et al., 2007) were asked to fill in a survey that sought to ascertain, amongst 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 that students did consider that this had been achieved. 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 specifically asked students to mark their agreement on a number of questions related to study skills (on a 5-point scale, 5 being 'strongly agree'). The questions related to the benefits related to time management and workload had a mean score of 3.11 (St. Dev. 0.96), note-making 3.03 (St. Dev. 1.02). The authors remarked that this was not an overly convincing result. A more general question, whether respondents thought SI had helped them to develop study and learning strategies, was more positive (3.72, St.Dev. 0.89). However, as the authors remark, the respondents may also have interpreted this to mean study skills directly related to content, rather than more general transferable academic skills.

One study where the effect of SI on academic skills development was studied by using a recognized instrument, was conducted by Ning and Downing (2010) in a Hong Kong university. They used the Learning and Study Strategies Inventory (LASSI) by Weinstein and Palmer (2002) to measure a range of skills and attitudes. It consists of ten factors clustered into three latent constructs, 'self-regulation' (e.g. time management), 'skills' (e.g. information processing) and 'will' (e.g. motivation). Ning and Downing collapsed the three latent constructs into one construct: 'learning competence'. They established that learning competence had a mediating effect on the learning achievement of both the SI participants and non SI participants. Comparison of SI participants and non SI participants revealed that of the ten different factors, for SI participants there was a significant effect for gains in information processing skills (d = 0.43 for SI participants versus d = 0.13 for non SI participants) and motivation levels (d = 0.39 versus d = 0.10). They conclude that academic achievement therefore is not just a function of learning competence, but that learning competence may 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 automatically result in enhanced 'well-being'. Some studies explicitly referred 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 included being part of a supportive environment, and being able to discuss difficult material with other students.

Although this was not always explicitly stated as such, being able to enroll in a voluntary academically-focused programme 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. The authors indicated that for the majority of the students in the focus group, this function of 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 significant 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 more 'intimidated' or more reluctant to ask questions in lectures. In the study by van der Meer and Scott (2009), of the 286 respondents (out of 345) who filled in the open-text question, 15% 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 was a 'support' or 'help' service, sometimes acted against students enrolling in SI, arguably often students who would most benefit from attending SI sessions; the originators of the program state "whether through denial, pride, or ignorance, students who need help the most are least likely to request it" (Martin & Arendale, 1993, p. 42). This was one of the motivations of Hodges, et al. (2001) to experiment with compulsory SI attendance for one group of students. They refer to studies that suggest a curvilinear pattern in help-seeking behavior, in other words, the least able are less likely to participate in SI.

Enhanced social relationships. Participation in SI or similar peer learning programmes, by its very nature enhances students' opportunity to meet other students, and potentially develop new friendships. Although not every SI programme explicitly states these social benefits in their design intentions, students' evaluation of SI typically seemed to yield comments that confirmed this benefit. Dobbie and Joyce (2008) conducted a number of focus groups with students who attended their peer learning programmes. 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 this 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) too suggests that introduction of SI in her English composition class has enhanced 'peer responses', however this is reported anecdotally. Similarly, Court and Molesworth (2008) reported that 11 of 33 survey respondents identified ‘social gains’ as a benefit of participation in SI, whilst 8 mentioned this 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, they reported that for the three semesters during which SI was offered, students chose not to attend tutorials. This was in spite of the faculty-organised tutorials being 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.

In 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 enabled 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 meant that effect sizes could not be calculated. The variability of methodologies, approaches, measures and variable characteristics (such as grade scales) 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 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 were carried out in a quasi-experimental study design. Some authors suggested that studying SI through an experimental design would enhance the value of the findings. Cheng and Walters (2009), although advocating for such an approach, sound a note of caution because of the great 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 did assign students to an experimental and control group.

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 to 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, Dochen and Joy (2001) sought to respond to this by experimenting with making SI compulsory.

Related to the question of non-participation 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 mode, the voluntary nature of participation. 'Problematic' meaning here that it makes it difficult to assess whether achievement gains resulted from the 'SI treatment' or resulted from self-selection into the programme, with the dominant assumption being that more motivated or academically able students may be more likely to participate. Bowles and Jones (2003-2004a) however, claim that the traditional approach to account for self-selection using a single equation Ordinary Least Square model (using pre-university entry scores as a dependent variable), underestimates the effect of SI. In their study they use a simultaneous equation - limited dependent variable model of SI effectiveness - and establish a larger effect for SI after controlling for academic ability. Academically less able students in their study participated in larger numbers in SI. They do 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 this. 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 around a third of our included articles, we did not find examples of sustained, theoretically-informed, rigorous adherence to an overarching qualitative methodology (for example, 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). While formal methods exist for detecting publication bias (such as the funnel plot or fail-safe n), they rely on formulaically comparing effect sizes. This has not been 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 Supplemental Instruction for the 2001-2010 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. This may be particularly useful for those considering introducing SI in their institutions or responding to those in their institutions who question the empirical evidence basis 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. Whereas in the field of learning support interventions there is a marked lack of evidence base in demonstrating effectiveness, SI has benefited from at least a degree of empirical research over more than forty years in many different countries.

Second, it 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: numbers of students involved in any group for which data is reported on, SI attendance requirements for membership in a particular group, the mean course grades and range of course grades (minimum, maximum) as well as standard deviations. Furthermore, where comparisons between different achievement data are provided, that significance levels (p values) 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 pre-entry 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 trans-national research collaboration projects to start to identify particular ways in which SI is implemented that seem to produce more convincing results than alternative ways. This would also benefit institutions who 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 under-researched. Effects of SI other than on academic achievement as expressed in course grades or course completion, would add to the value 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 under-represented minorities 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 programme 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 socio-economic status, first in family, and international students, may be of interest. As institutions world-wide become increasingly interested in tracking their "at risk" populations and analyzing the impact of different learning interventions, e.g. for quality and/or financial considerations, having rigorous research on the effectiveness of SI becomes increasingly important. Alongside this, the many institutions that consider factors that enable students to "thrive" in their transition into and through tertiary education, SI is one piece of the 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 US Department of Education is supported by published research from 2001-2010, and none are contradicted by research from this period. Importantly, however, none are 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 study 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.

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: 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, AZ: 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*. Minneapolis, MN: 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*(3), 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*(5), 459-464.

Bearman, M., & Dawson, P. (2013). Qualitative synthesis and systematic review in health professions education. *Medical Education, 47*(3), 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. *The Journal of Higher Education, 54*(1), 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*(2), 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*(4), 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*(30), 853.

Bradley, D., Noonan, P., Nugent, H., & Scales, B. (2008). Review of Australian Higher Education: Final Report. Canberra, ACT, 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 sciencies*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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*(1), 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*(6), 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, VIC, 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), 45-53B.

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*(7), 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: 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*(2), 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*(2), 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 2nd October, 2013, from http://www.umkc.edu/asm/si/si-docs/National Supplemental Instruction Report 98-03.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*(6), 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*(1), 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*(4), 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 *The Freshman Year Experience: Monograph Series* (2nd ed., Vol. 7). Columbia, SC: 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. In H. Levin (Ed.), *Alternatives to Developmental Education*. San Francisco: Jossey Bass.

McCarthy, A., Smuts, B., & Cosser, M. (1997). Assessing the effectiveness of supplemental instruction: A critique and a case study. *Studies in Higher Education, 22*(2), 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, NSW, 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*(8), 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*(3), 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*(4), 381-392. doi:10.1080/14703290903301784

Pawson, R. (2006). *Evidence-based policy: A realist perspective*. London: 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*(4), 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*(3), 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*(2), 353-387. doi:10.1037/a0026838

SI Staff from UMKC. (2005). *Supplemental instruction Supervisor Manual*. Kansas, MO: University of Missouri - Kansas City.

SI Staff from UMKC. (2010). Supplemental Instruction / Video Supplemental Instruction Annotated Bibliography Retrieved November 23, 2012, 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). Pubication bias: The Achilles Heel of systematic reviews? *British Journal of Educational Studies, 54*(1), 89-102. doi:10.1111/j.1467-8527.2006.00332.x

US Department of Education. (1992). Supplemental Instruction (SI): Improving Student Performance and Reducing Attrition. *Educational Programs That Work* (21st ed.). Washington, D.C.: US Department of Education.

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).

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.

Table 1

Overview of 29 included articles.

Study	**Field of study**	**Method**	**Study participants**	**Self- selection**	**Effectiveness**
(Bowles & Jones, 2003-2004a)	24 diverse disciplines	Quantitative: OLS vs simultaneous limited dependent variable model	3645	Pre-course GPA; Pre-tertiary achievement	Course grade
(Bowles & Jones, 2003-2004b)	24 diverse disciplines	Quantitative: Bivariate Probit model	3646	No	Retention
(Bowles et al., 2008)	Not stated	Quantitative: Two equation treatment effects	3905	Pre-tertiary achievement	Graduation

Study	Field of study	Method	Study participants	Self- selection	Effectiveness
(Bronstein, 2008)	Physical chemistry	Mixed methods: Case study	Not stated	Pre-tertiary achievement	Course grade; pass/fail; retention; wellbeing
(Cheng & Walters, 2009)	College Algebra and Probability; Pre-Calculus 1	Quantitative: Observational	534	Pre-tertiary achievement	Pass/fail
(Congos & Mack, 2005)	Chemistry	Mixed methods: Comparison between groups; Anecdotes	"thousands"	Pre-tertiary achievement	Course grade; pass/fail
(Court & Molesworth, 2008)	Creative media production	Quantitative: survey	34	Pre-tertiary achievement	Skills development; social; uncertainty reduction

Study	**Field of study**	**Method**	**Study participants**	**Self- selection**	**Effectiveness**
(Dancer et al., 2007)	Econometrics	Quantitative: Econometric models and descriptive statistics	890	Pre-course GPA; Pre-tertiary achievement	Course grade
(Dobbie & Joyce, 2008)	Accounting	Qualitative: Thematic analysis of focus groups	12	N/A	Academic skills; connectedness (student-student); capability (academic competence)

Study	**Field of study**	**Method**	**Study participants**	**Self- selection**	**Effectiveness**
(Fayowski & MacMillan, 2008)	First year calculus for non-majors	Quantitative: Quasi-experimental	869	Pre-course GPA; Pre-tertiary achievement; Comparison with prior offering that was not provided SI; pre-course GPA	Course grade; pass/fail
(Gattis, 2002)	Chemistry	Quantitative comparison; ANCOVA	142	Pre-tertiary GPA; pre-admission test; motivational control: expressed desire to attend	Course grade

Study	**Field of study**	**Method**	**Study participants**	**Self- selection**	**Effectiveness**
(Hafer, 2001)	English composition	Quantitative comparison	1500	Final grades comparison with previous offering	Course pass/fail; withdrawal
(Hensen & Shelley, 2003)	Biology; Chemistry; Mathematics; Physics	Quantitative: Comparison; ANOVA; ANCOVA; Chi-square	7339	Pre-tertiary achievement	Course grade; pass/fail
(Hodges & White, 2001)	History; Mathematics	Quantitative: experimental posttest-only control group	103	Pre-tertiary achievement	Mean semester GPA; comparison with tutoring
(Hodges et al., 2001)	History - US	Quantitative: ANOVA and post-hoc comparisons	432	Non-random assignment to groups; motivation scale	Course grade; pass/fail; Post-course GPA

Study	Field of study	Method	Study participants	Self- selection	Effectiveness
(Longfellow et al., 2008)	English composition	Mixed methods: Evaluative research study; survey; interview; descriptive statistics	216	Non-random 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

Study	Field of study	Method	Study participants	Self- selection	Effectiveness
(Miller et al., 2004)	Chemistry; Statistics	Mixed methods	1089 chemistry students; unclear number of statistics students	Pre-tertiary achievement	Final course marks
(Moore & LeDee, 2006)	Biology	Quantitative: Quasi-experimental	Not stated; we infer 1592 (398/semester over 4 semesters)	Pre-tertiary achievement	Course grade; pass/fail
(Ning & Downing, 2010)	Business	Quantitative: SEM; univariate analysis	430	Pre-tertiary achievement	Academic skills; GPA

Study	**Field of study**	**Method**	**Study participants**	**Self- selection**	**Effectiveness**
(Ogden et al., 2003)	Political science	Mixed methods	248	Pre-tertiary achievement; pre-course GPA	Course grade; post-course GPA; retention
(Parkinson, 2009)	Chemistry; Mathematics	Quantitative: quasi-experimental; ANOVA	66	Pre-tertiary achievement	Course grade; course pass/fail; task mark
(Peterfreund et al., 2008)	Chemistry; Biology; Statistics; Calculus	Quantitative	12423	Pre-tertiary achievement; pre-course GPA	Course grade; pass/fail; retention
(Phelps & Evans, 2006)	Mathematics (developmental)	Mixed methods	13800	A control group that was not offered SI	Course grade; pass/fail

Study	Field of study	Method	Study participants	Self- selection	Effectiveness
(Rath et al., 2007)	Biology	Quantitative	2698	Pre-tertiary achievement; Comparison with a group that was not offered SI	Course grade; pass/fail; retention
(Smith et al., 2007)	Surveying	Qualitative: Case study	332	Pre-tertiary 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

Study	**Field of study**	**Method**	**Study participants**	**Self- selection**	**Effectiveness**
(van der Meer & Scott, 2009)	Not stated	Mixed methods	345	Pre-SI task mark; interview	Academic skills; connectedness (student-student); capability (academic competence)
(Wright et al., 2002)	Mathematics (developmental)	Quantitative	Not stated	Pre-tertiary achievement	Course grade; pass/fail