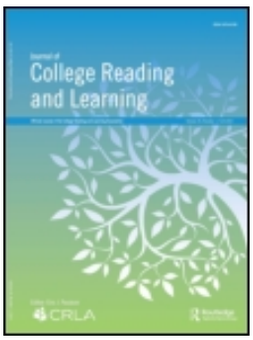


Routledge
Taylor & Francis Group

Journal of College Reading and Learning

ISSN: 1079-0195 (Print) 2332-7413 (Online) Journal homepage: https://www.tandfonline.com/loi/ucrl20

Evidence to Support Peer Tutoring Programs at the Undergraduate Level

Mitchell Colver & Trevor Fry

To cite this article: Mitchell Colver & Trevor Fry (2016) Evidence to Support Peer Tutoring Programs at the Undergraduate Level, Journal of College Reading and Learning, 46:1, 16-41, DOI: 10.1080/10790195.2015.1075446

To link to this article: https://doi.org/10.1080/10790195.2015.1075446

Published online: 05 Oct 2015.

Submit your article to this journal

Article views: 822

View related articles

View Crossmark data

Citing articles: 5 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ucrl20

Journal of College Reading and Learning, 46: 16–41, 2016
Copyright © College Reading and Learning Association
ISSN: 1079-0195 print/2332-7413 online
DOI: 10.1080/10790195.2015.1075446

Routledge
Taylor & Francis Group

Evidence to Support Peer Tutoring Programs at the Undergraduate Level

MITCHELL COLVER
Utah State University

TREVOR FRY
Florida Institute of Technology

The present study examined undergraduate peer tutoring in three phases. Phase I qualitatively surveyed students' perceptions about the effectiveness of tutoring. Phase II examined the usefulness of promoting regular use of services through a tutoring contract. Phase III utilized an archival, quasi-experimental approach to estimate the effect of peer tutoring on students who repeated an undergraduate course and either received tutoring or did not receive tutoring during their second attempt. Results revealed a significant estimated treatment effect, lending support to a causal relationship between tutoring and final course outcomes. Tutoring was especially beneficial for first-generation college students.

KEYWORDS academic support, first-generation college students, higher education, peer tutoring, program design, tutoring

Individualized tutoring has become a standard offering in the gamut of programs and services that institutions of higher education offer to help students persist toward graduation. Indeed, numerous articles have been written to provide empirical support for individualized tutoring at the undergraduate level, with a variety of methodologies being relied on to investigate the effectiveness of peer tutoring. A casual glance at the existing literature reveals a striking yet understandably low yield of empirically designed studies that support tutoring at the undergraduate level. In fact, a vast majority of the research that is available relies exclusively on correlational, qualitative, or other similarly limiting methodologies that make it difficult to glean insight into the causal impact that tutoring might have on student success.

Correspondence concerning this article should be addressed to Mitchell Colver, Utah State University, SOTS, 0122 Old Main Hill, Logan, UT 84322. E-mail: mitchell.colver@usu.edu

The literature review of articles examining one-on-one tutoring at the undergraduate level conducted for the current study resulted in only 29 studies that were ultimately considered to be of relevant value. This number is of particular importance because, as Holliday (2012) pointed out, "research about the effectiveness of tutoring has not kept pace with the widespread use of tutoring" (p. 21). Of these 29 articles, five used qualitative methodologies (American River College, 1993; Chen & Liu, 2011; Metcalfe, 1992; Mynard & Almarzouqi, 2006; Saunders, 1992), and 21 relied on quantitative methodologies, with three of these quantitative studies using methods that presented significant limitations to the meaningfulness of the data, such as using a small sample size (Comfort, 2011), using a low-impact intervention (Dioso-Henson, 2012), and presenting a self-described "pilot study" (Fuller & Denehy, 1972). An additional three articles offered valuable review or general discussion of issues related to tutoring literature (Holliday, 2012; Santee & Garavalia, 2006; Topping, 1996).

Of the 21 studies that used quantitative methodologies to evaluate the impact of peer tutoring on students' academic achievement, five demonstrated the effectiveness of tutoring without using a control condition, making it difficult to draw meaningful conclusions from the data (Bryer, 2012; Chaney, 2010; Topping, Watson, Jarvis, & Hill, 1996; Trevino & Eiland, 1980; Walker-Bartnick, Berger, & Kappelman, 1984). Ten relied on correlational methodologies that, by nature, cannot demonstrate causality (Cooper, 2010; Farmer, Lachter, Blaustein, & Cole, 1972; Hendriksen, Yang, Love, & Hall, 2005; Higgins, 2004; House & Wohlt, 1990; Lidren, Meier, & Brigham, 1991; Munley, Garvey, & McConnell, 2010; Reinheimer & McKenzie, 2011; Sobral, 2002; Walker & Dancy, 2007). Within these articles, selected results demonstrate that tutoring has a positive relationship with higher GPA and student retention.

Common Limitations to Effective Tutoring Research in Higher Education

An examination of the bulk of the quantitative articles reviewed reveals that many of the limitations demonstrated in the literature are not only common but understandable. Such limitations are even, at times, insurmountable given the nature of examining the effectiveness of an operational tutoring program in vivo. An exploration of some of the more common limitations follows.

Practical and Ethical Issues Related to Self-Selection and Random Assignment

Of the several fairly prevalent limitations found in the literature, one of the most common limitations is that students typically self-select to participate in tutoring. When the effectiveness of any tutoring program is being evaluated in vivo, true experimental design is virtually impossible because researchers are unable to engage in random selection and random assignment to control for individual differences such as the motivations that cause students to seek tutoring. This is a great challenge for researchers to overcome, because random assignment in actual tutoring scenarios can be unethical when certain participants are randomly assigned to receive tutoring while others are aware they are being denied access.

One solution to this problem is to use the methodology employed by Annis (1983), who performed research in a classroom setting but did so without using the actual curriculum and grades of the course as a measure for the effectiveness of tutoring. Instead, a mock lesson and associated nongraded assessment were used in place of the actual course content. In other words, Annis was conducting research using carefully designed measures in an experimental setting with remarkably high mundane realism, but without impacting students' final course grades. Although this type of research is nearly ideal, the need to regularly evaluate operational tutoring programs makes Annis' methodologies understandably less common in the literature.

Program Consistency and Empirical Continuity

Another limitation that arises in the literature with regard to the effectiveness of tutoring at the undergraduate level is that there is understandably low consistency among tutoring program models described in different research articles. Topping (1996) described more than seven different undergraduate-level models of tutoring that have been outlined in the extant literature. For example, some programs offer drop-in tutoring sessions, whereas others offer tutoring by appointment; some utilize a one-on-one format, and others offer small-group designs; some tutoring takes the form of reciprocal peer tutoring, and other tutoring occurs in a more structured, nonreciprocal tutoring model; certain programs employ faculty members or postgraduates as tutors, but many others rely on upper-level students (peer tutors) who have already demonstrated mastery of a particular subject. All of these differences make it difficult for one to achieve from the literature a clear picture of which models work and which models are not worth the investment of institutional resources. This cumbersome reality is compounded by the relatively few articles that examine tutoring in a manner that can demonstrate the causal impact of tutoring.

Correlation Does Not Equal Causation

Another common pitfall encountered by tutoring researchers, especially in light of self-selected access to most tutorial services, is a tendency to draw or imply causal conclusions from correlational data. For example, Cooper (2010) relies exclusively on correlational data while claiming that "tutoring appears to be an effective means of *improving* student retention and academic standing" (p. 33; emphasis added). This causal statement, even if tentative in voice ("appears to be") seems unusual in light of the author's own acknowledgment that "it is practically impossible to discern a causal relationship between tutoring and [some outcome] as the students using the [tutoring], and the number of times they use it, are self-selected" (p. 27). In other words, it is entirely possible that students who self-select to be tutored are characteristically above-average students or have greater motivational reserves than students who do not access tutoring. Rather than yield to such possible mediating factors, Cooper pursues two correlations between tutoring usage and subsequent cumulative GPA as a possible indicator of the effectiveness

of tutoring. However, these results still fall short of empirically demonstrating a causal impact of tutoring on meaningful student outcomes.

Sample Size and Appropriate Controls

Achieving a sizeable dataset is a challenging aspect of many types of social research and can be especially difficult for research related to tutoring because only a subset of students self-select to utilize the service. Even tutoring research that demonstrates some of the academic benefits of peer tutoring can be difficult to generalize to larger populations given relatively small sample sizes used for analysis (e.g., Comfort, 2011; Fuller & Denehy, 1972). Changes in coordinator staffing, upgrades to data tracking systems, or lack of longitudinal tenacity on the part of the researcher can easily interfere with consistent, contiguous tracking of program data. This makes it difficult to assess programs effectively and to transform the results into meaningful, publishable research. As Holliday (2012) explained, "Learning center staff members may not have the personnel or time to undertake a demanding study, or they may not have the training or interest in research methodologies or statistical analysis" (p. 22). In the context of the present study, for two and a half years the dataset failed to approach a sample size that would provide sufficient statistical power. In other words, the current results were made possible only by tracking data consistently and uniformly across three years of the program.

Other sampling issues include the lack of an appropriate control condition for evaluating the effectiveness of tutoring. Several of the articles reviewed for the present study demonstrated an inability to use appropriate controls for comparison (Bryer, 2012; Muraskin, 1997; Sobral, 2002; Walker-Bartnick et al., 1984). For example, although some studies did not include any type of control group, others took an archival approach and obtained academic achievement data from past students, identifying those students who were not on file as having received formal peer tutoring (Sobral, 2002). The limitation of this method was that Sobral does not appear to have required students in the control group to have taken the same courses or to have enrolled during the same academic terms as the experimental participants. One way to control for such extraneous variables is to identify sets of students (participants and corresponding controls) who have attended the same sections of the same courses during the same academic term. This approach allows the researcher to scrutinize the effects of peer tutoring on course outcomes by comparing groups of students whose experiences presumably differ only in that some students received tutoring and others did not. Certain ethical considerations related to the use of this approach are discussed in greater detail following, especially in terms of the work of Arco-Tirado, Fernández-Martin, and Fernández-Balboa (2011).

Three Exceptional Studies

Three articles (Annis, 1983; Arco-Tirado, Fernández-Martín, & Fernández-Balboa, 2011; Lake, 1999) are of particular note given their unique quantitative

methodologies. A quantitative study by Arco-Tirado et al. used random assignment to evaluate the impact of a peer tutoring program for students at the University of Granada in Spain. Freshman students were randomly assigned to one of two groups: (a) those being tutored by upper-class peers or (b) those not being tutored at all. At the conclusion of the peer tutoring program for the selected term, researchers compared the groups' GPA scores. Although the analysis did not reveal statistically significant differences, the design of this study was sufficient to have potentially demonstrated a causal impact of tutoring.

There are certain ethical concerns related to randomly assigning particular students to receive tutoring while withholding the service from students who are randomly assigned to a control group. This is especially true in a situation when actual grades are on the line. As explained by Kimmel (1988),

> When some direct loss to untreated controls can be foreseen [by the controls] prior to the study, their inclusion may be partly indefensible. This issue is potentially complicated in cases where failing to receive a benefit is experienced as a loss by those affected . . . [especially] if untreated controls become disappointed about their relative deprivation. (p. 80)

Because the participants of the Arco-Tirado et al. (2011) study were apprised of their assignment to an untreated control condition, unaccounted-for variability makes interpreting the outcomes of this study somewhat difficult.

Another quantitative study of particular note, conducted by Annis (1983), demonstrated positive effects of peer tutoring while using random assignment but did so in a more ethical manner than did Arco-Tirado et al. (2011). The study showed a positive impact of tutoring with respect to the six levels of Bloom's Taxonomy of Educational Objectives (knowledge, comprehension, application, analysis, synthesis, and evaluation; Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956) on the assigned reading of an article in a world history course. Students were randomly assigned to one of five conditions: (a) read the lesson, (b) read the lesson with the intent to tutor the lesson, (c) read the lesson with the intent to tutor the lesson and then actually tutor the lesson, (d) receive tutoring on the lesson without reading, and (e) read the lesson and receive tutoring on the lesson. This methodology provides a sound framework for between-subjects comparisons. The methodological advantage of this approach is that the article for the assigned reading does not appear to have been part of the students' graded work for the course, and untreated controls remained unaware of being deprived of the tutoring intervention. Results showed that students who read the lesson and provided tutoring to other students demonstrated the highest exam scores. Additionally, students who read and received peer tutoring on the lesson had significantly higher scores on items pertaining to knowledge and application than students who only read or only tutored. Furthermore, students who received tutoring scored significantly higher than students in all other conditions on items related to synthesis. These results not only highlight the academic benefits of tutoring others but also support the idea that students who are actively involved in self-study while receiving peer tutoring have an academic advantage compared to those who do not.

A third quantitative article of particular note (Lake, 1999) found that students who received peer tutoring in an advanced physiology course demonstrated significantly lower decline in course grades (compared to a previous semester of the same subject) than students not involved in peer tutoring. Researchers measured declines in grades from one semester of introductory physiology to a following semester of advanced physiology for groups of students who did and those who did not receive peer tutoring during the advanced course. Although grades declined in both groups for the advanced course, students who received tutoring had significantly lower declines in course outcomes compared to the outcomes of the introductory course. The design allowed Lake to evaluate the effect of peer tutoring on tutored students in comparison with a control group of students who did not receive peer tutoring. This design is highly useful, because it allows for a baseline comparison between an experimental group and a control group. Limitations of Lake's (1999) research are a relatively small sample size and the decision to use the introductory course as a baseline for comparison against a following semester of an advanced course. Additionally, Lake did not demand that controls be from the same course sections during the same academic terms as experimental participants, which introduces the possibility of confounding variables related to instructional style and other varying elements of the course experience.

THE NEED FOR FURTHER EMPIRICAL ANALYSIS

Despite successes reported in past research, there continues to be a somewhat understandable lack of empirical, causal support for using individualized tutoring at the undergraduate level. This circumstance is unfortunate in a time of rapidly shifting financial climates and a need to emphasize retention and completion over enrollment (National Conference of State Legislatures, 2014). Indeed, as explained by Lidren et al. (1991), the striking lack of "conclusive evidence to provide rationale for the widespread implementation of effective peer tutoring programs in college settings" (p. 70) is cause for concern.

In direct response to these issues and in an effort to contribute further qualitative and quantitative support for the effectiveness of peer tutoring, the present study sought to follow the recommendation of Topping (1996): "It is essential that subsequent research strives to achieve adequate quality in design and execution, preferably including control groups or comparison groups which are truly comparable, and addresses issues of achievement gain and parameters of successful course completion" (p. 339). The purpose of the present study was to examine, in an empirically robust manner, the contribution of peer tutoring to students' grades in three separate phases of methodologies. The third and final phase is unique in the field in that it uses a nonequivalent comparison group design as a way to mitigate the influence that self-selection to be tutored can have on research outcomes. Specifically, as described by May (2012), the nonequivalent comparison group design allows researchers "to *estimate* the treatment's effect in the absence of random assignment" (p. 489; emphasis added). This feature of the design is particularly useful to programs that must evaluate academic services that are provided to students who are allowed to self-select to participate in interventions such

as open-enrollment tutoring. Thus, the nonequivalent comparison group design, although not a true measure of the causal impact that a program can have, is the next best thing to a situation that allows for random assignment to occur.

PROGRAM ORGANIZATION

Given the variety of program models available for implementing tutoring at the undergraduate level (see Topping, 1996), a thorough description of the model used for the present study is in order. The academic services program investigated in the present study was initially implemented at Eastern Washington University to provide group learning services similar in nature to the Supplemental Instruction (SI) program designed by staff at the University of Missouri–Kansas City. Although these peer-facilitated study groups were the primary offering of the program for many years, in 2010 it was decided that individualized tutoring, which had previously been available only to students participating in a federal TRiO Student Support Services program, would be expanded to be universally available to all students on campus. Funding was allocated, and the program was established to provide individualized tutoring to as many students as was financially feasible. Not surprisingly, before the end of the inaugural academic term of the newly created individualized tutoring program, demand for the service had increased to an unmanageable level. This warranted the creation of a protocol that would funnel students to the services most appropriate for their individual needs. Specifically, it was decided that unless students were part of an exceptional category (for example, those retaking the course or those unable to attend the scheduled study group for any given course), peer-facilitated study groups would remain the preferred service to meet students' academic support needs instead of individualized tutoring services.

Any student interested in participating in individualized tutoring was scheduled to meet with the tutor coordinator for a 15-minute orientation meeting. This meeting was designed to qualify the students' needs and to ensure that each student was directed to the most appropriate service for his or her unique circumstances. Students eligible to participate in individualized tutoring had to meet only one of the following seven qualifications:

- The student is retaking the course.
- The student has a registered disability.
- The student is a veteran or member of the Armed Services.
- The student is working with a special program advisor who has provided a professional recommendation that the student receive tutoring.
- The student is enrolled in a course for which no peer-facilitated study group is offered.
- The student is enrolled in a course for which a peer-facilitated study group *is* offered, but the student is unable to attend due to a course scheduling conflict.
- The student has received individualized tutoring from the program during a previous quarter and demonstrated responsible use of the service (the "grandfather clause").

As in the study by Lidren et al. (1991), unless students fell into one of the above seven categories, peer-facilitated study groups remained the preferred service available for enhancing their academic performance. The work of Lidren et al. serves as a noteworthy guide for using random assignment and sufficient controls to assess the effectiveness of individualized tutoring compared with group tutoring for self-selected participants. Their results demonstrated that those who received either type of tutoring (individualized or group) achieved significantly higher grades than did controls, or students who did not participate in weekly tutoring sessions. However, no significant differences emerged between the experimental groups (individualized vs. group tutoring). In other words, although one might expect individualized tutoring to influence student outcomes to a greater degree, Lidren et al. found that both individualized and group tutoring seemed to influence student outcomes to a similarly effective extent. As such, offering individualized and group tutoring programs simultaneously can be a cost-effective, two-pronged approach to meeting academic support needs of undergraduate students. The decision to funnel the vast majority of students into peer-facilitated study groups not only respected financial advantages but also honored the effectiveness of group study sessions and helped reserve individualized tutoring for those students who demonstrated the greatest academic need (as outlined in the seven qualifying categories listed above). It is also worth noting that tutoring services for English and writing courses were not provided through this program but instead were offered exclusively by the institution's Writers' Center.

For students who were found during the tutoring orientation meeting to be eligible to receive individualized tutoring, a contract was established to ensure their commitment to use the program effectively and on a regular basis. This contract includes the following five student commitments:

- I agree to meet with my tutor for the duration of the academic term for at least two 50-minute sessions each week for the duration of the quarter (finals week optional).
- I agree to schedule my appointments in advance and ensure that the sessions are regularly occurring similar to a class (for example, every Monday, Wednesday, and Friday from 1:00 to 1:50 p.m.).
- I agree to achieve or consistently strive to achieve at least 15 sessions of tutoring throughout the duration of the academic term.
- I agree to notify my tutor at least 12 hours before canceling a scheduled session. In case of an emergency, I agree to notify my tutor or the Tutor Coordinator as soon as possible.
- I agree not to accrue more than two "no-shows" during the academic term. (A no-show constitutes a scheduled session that I do not cancel with 12+ hours of notice and then do not attend.)

The purpose of establishing and maintaining this contract with each student was to ensure that students benefited from "greater student ownership of the learning process," (Topping, 1996, p. 325), which is one of the advantages of receiving tutoring as described by Topping. Additionally, the model was specifically designed to encourage regularly occurring distributed practice, one of the

most highly effective study strategies evaluated by Dunslosky, Rawson, Marsh, Nathan, and Willingham (2013). Each of the five requirements listed above was crafted with particular attention to enabling students to engage the tutoring process early and often. Thus, meeting these requirements would result in the greatest amount of distributed practice throughout the duration of the quarter. When students eligible to receive individualized tutoring agreed to the above requirements, they were assigned to meet with a specific peer tutor for the duration of the quarter. Students unwilling or unable to meet the contract or found to be ineligible for individualized tutoring, based on the eligibility requirements previously described, were referred to attend the offered peer-facilitated study groups.

Because the program described in this study is certified by International Tutor Training Program Certification (ITTPC) of the College Reading and Learning Association (CRLA), peer tutors were selected and interviewed on criteria requiring them to have received at least a 3.0 in the courses they tutored and to receive a recommendation from a content-specific faculty member or from the tutor coordinator (College Reading and Learning Association, 2015). CRLA's ITTPC also requires tutors to participate in ongoing tutor training and to participate in a formal evaluation process. Using this model of highly structured training and evaluation is supported by the work of Cohen, Kulik, and Kulik (1982), who demonstrated that tutor training leads to greater effectiveness of tutoring services in general and that a highly structured tutoring model is associated with the effectiveness of the tutoring.

At the end of each academic term, electronic qualitative surveys were sent to all students to allow them to evaluate the program and their perception about the impact the tutoring had on their well-being as a student and on their grades. Tutoring session data were collected and analyzed to determine whether each student had met the contract. Students who accrued more than two no-shows, more than three weeks in which they fell below the required two-sessions-a-week minimum, or more than one week in which they did not meet with their tutor at all (Thanksgiving week excluded) were classified as having failed to meet the contract. In other words, students whose attendance was not consistent and frequent throughout the entire quarter did not classify as having met the contract. Grades and other important data were solicited from institutional research and, once compiled, were curated for accuracy.

PHASE I

Phase I of the present study was a preliminary, qualitative assessment of students' perceptions about the relative effectiveness of the tutoring program. Effectiveness was assessed in terms of perceptions about the impact tutoring had on general student well-being, important course objectives, and final course grades.

Method

The use of this kind of preliminary, qualitative methodology can help establish that, at least in the minds of those served, the program is carrying out the ends

for which it was originally designed. Results achieved in Phase I gave impetus to continuing to pursue more complex methodologies that would assess the effectiveness of tutoring on a quantitative level, leading to the creation of Phase II and Phase III.

Participants

Participants for Phase I included 349 unique students who utilized individualized tutoring services during the 2012 calendar year. Many of these students were tutored for more than one course or during more than one quarter, resulting in 520 individual instances of student/tutor pairings that were used for analysis.

Data Collection

At the end of each academic term (Winter, Spring, and Fall quarters), a brief electronic program evaluation was sent to each student who had been tutored. The first set of questions asked the student to rate the impact of tutoring relative to ten statements with which the student indicated their level of agreement or disagreement on a 5-point Likert-type scale (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree). The tenth statement in this series, "Tutoring changed my life," was meant to test the sensitivity of the questionnaire, as levels of agreement for that statement would likely be realistically far lower than more relevant questions related to course outcomes. One final question asked the students to estimate the impact that tutoring had on their final course grade, with options ranging from "Lowered my grade by 1.0 grade points or more" to "Increased my grade by 1.0 grade points or more," assessed at 0.2 intervals throughout this range. The survey was completed anonymously and, in compliance with the Family Educational Rights and Privacy Act (FERPA) regulations, submitted to a secure server within the direct control of the institution. Of the 520 possible responses, 142 surveys were returned, representing a 27% response rate.

Results

As shown in Table 1, responses represent the generally positive perceptions of students about the effectiveness of the tutoring program, assessed on multiple factors. In the first section of qualitative responses, students indicated their strongest level of agreement with the statements, "Tutoring improved my understanding of the course material" ($M = 4.33$, $SD = .94$) and "Tutoring helped me do better on assignments" ($M = 4.27$, $SD = .91$). These two statements, along with "Tutoring helped me to be a more confident student" ($M = 4.14$, $SD = .97$), also represent the greatest consistency of response, having the lowest standard deviations of all the statements assessed. In line with the hypothesis that the statement, "Tutoring changed my life," would be least likely to be agreed with, responses represented the lowest level of agreement ($M = 3.24$, $SD = 1.20$), indicating that, in general, students were likely cognizant of the answers they were submitting. Furthermore, a plurality of students (38%) indicated that tutoring impacted their final course grade

TABLE 1. Results of a Qualitative Tutoring Program Evaluation

Section I: Assessed on a level of 1 (strongly disagree) to 5 (strongly agree)	*M*	*SD*
1. Tutoring helped me get a better grade.	4.15	1.02
2. Tutoring helped me do better on tests.	4.11	1.03
3. Tutoring helped me do better on assignments.	4.27	.91
4. Tutoring helped me do better reading the textbook.	3.8	.99
5. Tutoring helped me to be a more confident student.	4.14	.97
6. Tutoring improved my study skills.	4	1.01
7. Tutoring improved my understanding of the course material.	4.33	.94
8. Tutoring helped me stay in school.	3.52	1.24
9. Tutoring helped me through significant challenges.	3.95	1.05
10. Tutoring changed my life.	3.24	1.20
Section II: Students indicated the estimated impact tutoring had on course grade	# of Responses	% of Responses
Increased my grade by 1.0 grade points or more	54	38%
Increased my grade by .8 to 1.0 grade points	12	8.5%
Increased my grade by .6 to .8 grade points	14	9.9%
Increased my grade by .4 to .6 grade points	17	12%
Increased my grade by .2 to .4 grade points	15	10.6%
Increased my grade by .0 to .2 grade points	13	9.2%
Tutoring had no effect on my grade (0 point difference)	12	8.5%
Lowered my grade by .0 to .2 grade points	2	1.4%
Lowered my grade by .2 to .4 grade points	0	0%
Lowered my grade by .4 to .6 grade points	0	0%
Lowered my grade by .6 to .8 grade points	0	0%
Lowered my grade by .8 to .1.0 grade points	0	0%
Lowered my grade by 1.0 grade points or more	3	2.1%
Totals	**142**	**100%**

by 1.0 grade points or more, with 88.2% of respondents indicating that tutoring improved their grade at least to some degree.

Discussion

Results of Phase I indicate that, for the tutoring program in question, students found the services to be generally effective. Responses show that tutoring was most consistently perceived to impact understanding of the course material, to improve work on assignments, and to increase confidence in students, in addition to influencing other meaningful outcomes, as shown in Table 1. Additionally, 88.2% of respondents indicated that tutoring had improved their grades at least to some degree. In general, these results reveal that the tutoring program was, at least in the minds of those served, impacting the customary academic outcomes that tutoring programs generally seek to address. As such, further quantitative investigations seemed appropriate as outlined in Phase II and Phase III.

PHASE II

Phase II was conducted to evaluate the general effectiveness of an operational tutoring program from a preliminary, correlational standpoint. The use of preliminary, correlational methodologies can help establish whether meaningful relationships exist between participation in a tutoring program and the academic outcomes that tutoring seeks to influence.

Method

Phase II sought to investigate the relationships between various factors (such as total hours of tutoring, week of onset of tutoring, and number of active weeks in tutoring) and the final grades that students received for their courses. Additionally, because so many different program models exist in the realm of undergraduate tutoring, Phase II also sought to investigate the effectiveness of the specific program model in question. Results achieved in Phase II were used to provide impetus for conducting Phase III, which necessarily used more advanced methodologies.

Participants

Participants for Phase II included 661 unique individuals who utilized individualized tutoring services during a three-year period from January 2011 to December 2013. Many of these 661 students were tutored for more than one course or during more than one quarter, resulting in 1,113 individual instances of student/tutor pairings that were used for analysis. All students included in Phase II of the study completed the courses for which they received tutoring. Additionally, students who received tutoring for preuniversity requirements (such as precollege algebra), as well as students classified as having already completed one bachelor's degree (postbaccalaureate and graduate students) were excluded from this phase of the study. Proportions of students from each class level were approximately as follows: 19% freshman, 25% sophomore, 37% junior, and 19% senior. Additionally, the most common academic disciplines for which students received tutoring, as observed in our sample, were chemistry and math, with Chemistry for the Health Sciences (19.8%), General Chemistry (13.5%), Pre-Calculus (7.5%), and Mathematical Reasoning (6.1%) being the most commonly tutored specific courses.

Data Analysis

Because the tutoring program in question was structured to use a contract to encourage students to engage in regularly occurring distributed practice (with emphasis on beginning tutoring early in the academic term), several correlations were performed to determine the relationship between final course grades and a handful of other factors, including number of weeks active in tutoring, total hours of tutoring received, how early the student sought tutoring (week of onset), whether or not the student met the contract, and student cumulative GPA. Additionally, because the data for course grades were not found to be normally

distributed, a Kruskal–Wallis test was run to determine if students who met the contract demonstrated significantly higher grades than students who did not meet the contract.

Results

The results of the correlations performed revealed a significantly positive relationship between students' course grades and a variety of important variables: number of weeks active in tutoring, $r(1113) = .21$, $p < .001$; total hours of tutoring received, $r(1113) = .18$, $p < .001$; how early the student sought tutoring (week of onset), $r(1113) = -.15$, $p < .001$; whether or not the student met the contract, $r(1113) = .14$, $p < .001$; and student cumulative GPA, $r(1113) = .47$, $p < .001$. Because GPA represented the strongest relationship with overall course grades, a second set of correlations was run on these same factors while controlling for cumulative GPA. Results remained consistent and significant, with only a slight drop in r-values: number of weeks active in tutoring, $r(1105) = .17$, $p < .001$; total hours of tutoring received, $r(1105) = .15$, $p < .001$; how early the student sought tutoring (week of onset), $r(1105) = -.10$, $p < .001$; and whether or not the student met the contract, $r(1105) = .12$, $p < .001$.

In order to investigate if a significant difference in final course grade existed between students who chose to adhere to the tutoring contract and students who did not, a Kruskal–Wallis test was conducted following the results of a Kolmogorov–Smirnov normality check that demonstrated that the data were not normally distributed. Of the 1,113 instances of tutoring included in this sample, 853 instances (76.6%) represent students who met the tutoring contract, with 260 instances (23.4%) representing students who failed to meet the tutoring contract even though they had signed up for tutoring and agreed to adhere to the requirements. Results of the Kruskal–Wallis test showed that students who fulfilled the tutoring contract received significantly higher grades in the course for which they received tutoring ($M = 2.46$, $SD = 1.07$) than students who received tutoring but failed to meet the contract ($M = 2.09$, $SD = 1.28$), $H(1) = 13.646$, $p < 0.001$, with a mean rank of 576.63 for students who met the contract and a mean rank of 492.59 for students who did not.

Discussion

The correlations found to exist between course grade and variables related to the contract guidelines were found to be significant yet yielded relatively low r-values. This finding is not entirely discouraging, because a wide array of variables probably contribute to final course grades. However, given the large sample size, the significance of these correlations is meaningful. The relationship between final course grade and various factors related to the contract, especially when controlling for cumulative GPA, seems to indicate that requiring students to meet a tutoring contract in order to receive tutoring services might be an influential strategy.

Interestingly, and perhaps most notably, the number of weeks students were active in tutoring constituted the strongest relationship with final course grades

(.21), even over the total hours of tutoring students received (.18). This finding could point to the importance of encouraging students to distribute tutoring sessions throughout the academic term rather than rushing to fit them in all at one time. As has been shown in past research, compacting learning into single episodes of time significantly impairs the learning process, especially in comparison to learning episodes that are separated by even short intervals (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). As was shown in the work of Dunlosky et al. (2013), distributed practice is one of the most highly effective study strategies available to students. Therefore, having to attend tutoring on a regular, ongoing basis could more than likely be a particularly important factor related to the effectiveness of tutoring.

The results of the Kruskal–Wallis test conducted in Phase II, which showed significantly higher course grades for students who chose to adhere to the tutoring contract over students who did not, indicate that inviting students to meet some kind of tutoring contract is at least one effective program model available to institutions of higher education. Although this result is encouraging, it is important to note that it does not demonstrate a causal relationship (given that students self-select whether to meet the contract). However, the degree to which students' grades differ between the two groups (2.46 to 2.09) is remarkable and does provide impetus for further investigation. In sum, the results of Phase II serve to point toward the importance of distributed practice and, in general, the importance of encouraging students to meet some kind of tutoring contract in order to receive services. These results also suggest a need for further investigation, as outlined in Phase III.

PHASE III

Phase III sought to expand on Arco-Tirado et al. (2011), Annis (1983), and Lake (1999) by using an archival approach to gather repeated-measures and paired data from students who received tutoring and those who did not. This approach, which included data across 3 years, yielded a large enough sample, including controls, to evaluate the effects of peer tutoring across a variety of college courses.

Method

In contrast to Arco-Tirado et al. (2011), a primary goal of Phase III within the present study was to evaluate the effect of peer tutoring on students' course grades without subjecting any individual to academic disadvantage through random assignment to an untreated control condition. Similarly, Phase III within the present study sought to expand on Annis' (1983) design by measuring changes in actual course outcomes (grades) between students who received tutoring and their classmates who did not, rather than measuring the effects of tutoring on a single, nongraded examination. Additionally, Phase III sought to expand on Lake's (1999) methodology by identifying students who repeated a course and happened to receive tutoring during their second attempt in the same course, rather than an advanced version of the same subject. Using a nonequivalent comparison group

design, Phase III also served to estimate, as accurately as possible (in the absence of random assignment), the causal impact that tutoring had on final course outcomes.

Data Collection

Data for Phase III were collected by examining 3 years of program rosters to identify students who would appropriately serve as experimental participants. Specifically, in order to qualify as experimental participants, students had to have taken any one course on two separate occasions, the second occasion being a repeat attempt. The students received tutoring only during their second attempt to finish the course. Additionally, in order to qualify, these students had to have met the tutoring contract described previously, during the academic term for which they were receiving tutoring. Once a cohort of qualifying students was established, course records were pulled from the institution's database in order to identify appropriate controls for each experimental participant. An appropriate control was defined as any individual who, simply by chance, had taken and repeated the same courses as their experimental counterpart, in the same sections with the same instructors during the same years and academic terms *for both their first and second attempts*. Additionally, in order to qualify for the control condition, individuals could not have received tutoring through the program under study during either of their attempts. Interestingly, multiple controls were found for many of the participants. This design is graphically demonstrated in Table 2.

This design is known as a nonequivalent comparison group design and is noteworthy because it involves participants who self-select to participate in the treatment condition; in this case, the students choosing to receive tutoring during their second attempt to successfully complete a course. Parametric limitations are known to exist for this type of quasi-experimental design and are discussed in greater detail in the section on limitations. The advantage of a nonequivalent comparison group design is that it improves on the methodology of Lake (1999), who used two separate levels of a physiology course (first semester and second semester) for the pretest/posttest measures. This is somewhat problematic because using two different sets of course content means that a true baseline is not necessarily established. Additionally, Lake did not include participants who had experienced the same course during the same year with the same instructor in either the experimental or control conditions; instead, Lake drew on course data from varying academic terms to constitute the study's sample. The design of Phase III of the present study ensures that the pretest and posttest measures were drawn from the same course for each set of participants. The use of appropriate controls

TABLE 2. Did the Student Receive Individualized Tutoring from the Program in Question

	. . . during their 1st attempt?	. . . during their 2nd attempt?
Experimental Cohort	No	Yes
Control Cohort	No	No

and an accurate baseline not only make Phase III unique in design but also allow for the treatment's effect to be estimated in the absence of random assignment.

Participants

A total of 136 students were included for analysis in Phase III, with 46 students classified as experimental participants (received tutoring only during their second attempt) and 90 students identified as appropriate for the control group (matched to experimental participants and having received no tutoring during either attempt). Participants included 91 females and 45 males, with roughly 75% being first-generation college students. Data were compiled over the course of three years, from January 2011 to December 2013. The average interval between the first and second attempt was roughly two academic quarters ($M = 2.06$, $SD = 1.26$), with 50% of the participants repeating the course for the second time during the quarter immediately following their first attempt. The most common academic disciplines observed in the sample were chemistry and math, with Chemistry for the Health Sciences (29.4%), Organic Chemistry (25%), and General Chemistry (22%) being the most common courses included in this phase of the study.

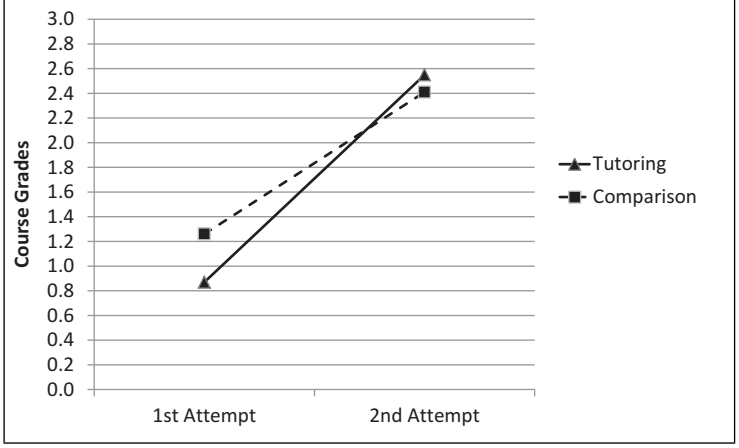
Data Analysis

A nonequivalent comparison group design was used to estimate the effect of tutoring on students' grades during their second attempt. As stated, the sample included undergraduates who repeated a course while receiving tutoring only during their second attempt and counterpart controls who repeated the same class but did so without receiving tutoring during either their first or second attempt. Initially, a repeated-measures ANOVA was used to compare grade changes between first and second attempts for all participants in order to determine if tutoring was influencing final grades. Subsequent to this analysis, an ANCOVA controlling for cumulative GPA was run using grades for the first attempt as a pretest covariate in order to more fully outline the effect size of the tutoring intervention. Additionally, exploratory analyses related to first-generation status were conducted.

Results

To investigate whether or not tutoring causally impacted students' course grades, a repeated-measures ANOVA was performed to examine the effects of tutoring on grade changes between first and second course attempts among college students repeating particular courses. There was a significant main effect for grade change, $F(1, 134) = 240.51$, $p = .000$, $\eta^2 = .64$, revealing that grades generally improved from all participants' first attempt ($M = 1.13$, $SD = .96$) to their second attempt ($M = 2.46$, $SD = 1.07$), regardless of whether tutoring was received. More importantly, we also observed a significant crossover interaction between changes in course grades and the presence of tutoring, $F(1, 134) = 8.8$, $p = .004$, $\eta^2 = .06$ (see Figure 1), indicating that receiving tutoring during the second attempt was

FIGURE 1. Changes in course grades from first to second attempts between groups.



associated with a higher increase in grade than repeating the course without tutoring. In other words, students who received tutoring demonstrated a higher mean increase in grade, compared to controls, from their first final grade ($M = .87$, $SD = .83$) to their second final grade ($M = 2.55$, $SD = .90$) during a repeated attempt. This increase was significantly greater than the increase in grades for students who did not receive tutoring in either their first attempt ($M = 1.26$, $SD = 1$) or second attempt ($M = 2.41$, $SD = 1.14$) for the same courses.

As described by May (2012), a parametric limitation to using a nonequivalent comparison group design is that the participants are allowed to self-select to participate in the treatment condition, possibly resulting in a false representation of the treatment effect within the data. Typically, in order to account for this parametric confound, an adjusted pretest score (in this case, the grade for first attempt) is calculated based on the reliability of the pretest measure (May, 2012). However, because a reliability statistic cannot be computed for a single course grade, pretest data were adjusted using an extremely conservative reliability level (.01), which served to set an unusually high bar for achieving statistical significance. In other words, whereas a rather poor reliability level would be as low as perhaps 50% or 30%, the authors of the present study chose to conjecture a reliability level of 1%, effectively raising the bar for demonstrating an effect of the treatment. When using this extremely low hypothetical reliability level to adjust the grades of the first attempt according to the methods of May (2012), results still remained consistent with the original analysis, with tutoring having a significant effect on grades during the second attempt, $F\ (1, 134) = 7.91$, $p = .006$, $\eta^2 = .056$. Therefore, although a true reliability adjustment could not be made, this approach stacked the deck, so to speak, in favor of the null hypothesis and sufficiently ensured that self-selection was not reasonably likely to have produced a false representation of the treatment effect within the data.

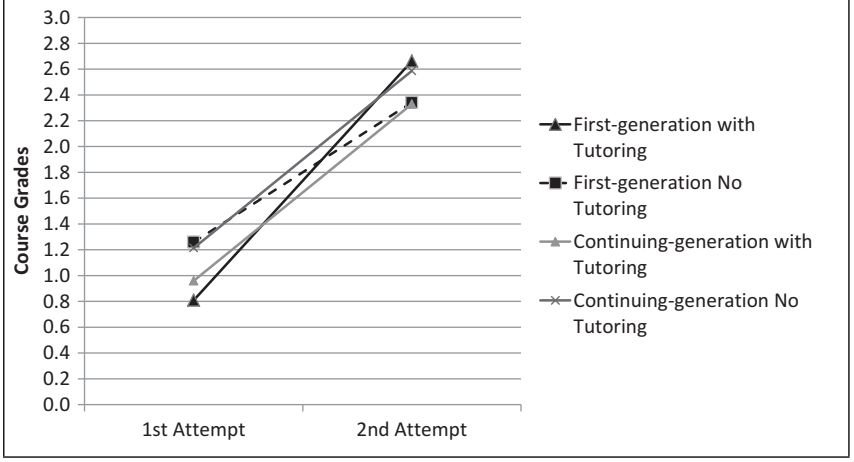
Because the effect of tutoring was found to be significant using a repeated-measures ANOVA, an ANCOVA was performed using the pretest grades as a covariate in order to more fully analyze the size of the effect that receiving tutoring had on students' second attempts. As explained by Dugard and Todman (1995), using the pretest as a covariate eliminates a great deal of statistical noise present in a repeated-measures design and allows for a more accurate estimate of the treatment effect size. Once again, a significant relationship between the presence of tutoring and successful course outcomes was found to exist, $F(1, 133) = 4.75$, $p = .031$, $\eta^2 = .034$, with a reduction in the effect size compared to the previous model. Students who received tutoring demonstrated significantly higher estimated course grades in their second attempt when adjusting for grades at first attempt ($M = 2.70$) than the estimated course grades of students who did not receive tutoring ($M = 2.33$). As one can see, this model significantly increased the difference in performance between the experimental participants and controls during the second attempt. We ran the same ANCOVA including GPA on the first attempt as another covariate to control for overall academic quality of each student, in order to factor out academic proficiency as a possible confounding variable. For example, having greater academic proficiency could cause students to seek and utilize resources such as tutoring more ardently than students with lower academic proficiency. Again, when using first-attempt grades and GPA as covariates, a significant effect of tutoring emerged, $F(1, 132) = 4.77$, $p = .031$, $\eta^2 = .035$. Once again, to address the parametric limitation to using a nonequivalent comparison group design (May, 2012), grades at first attempt were adjusted using the extremely low hypothetical reliability level (.01), and the results of the ANCOVA showed a much larger effect for tutoring on grades during second attempt, $F\ (1, 132) = 23.6$, $p < .001$, $\eta^2 = .152$. This increase in effect size demonstrates that, when reducing noise introduced to the data by variation in the pretest covariate, a meaningful relationship between tutoring during the second attempt and successful course outcomes emerges.

Next, we ran a 2 (tutoring) $\times$ 2 (first-generation status) factorial ANCOVA looking at the effects of peer tutoring on course outcomes between first-generation students and continuing-generation students after controlling for course grades and cumulative GPA at first attempt. Five participants in our sample were excluded from the analysis because data on first-generation status was not available. A significant crossover interaction emerged between tutoring and first-generation status, $F(1, 125) = 4.37$, $p = .039$, $\eta^2 = .034$ (see Figure 2). First-generation students who received tutoring had higher second-attempt grades ($M = 2.66$, $SD = .91$) than first-generation students who did not receive tutoring ($M = 2.34$, $SD = 1.25$), whereas continuing-generation students who received tutoring had lower second-attempt grades ($M = 2.33$, $SD = .95$) than continuing-generation students who did not receive tutoring ($M = 2.59$, $SD = .79$). These results suggest that peer tutoring is of particular benefit to individuals identified as first-generation college students.

Discussion

The purpose of Phase III of this study was to estimate the treatment effect of tutoring on course grades among students repeating an undergraduate course in the

FIGURE 2. Changes in course grades from first to second attempts between first-generation and continuing-generation students who did and did not receive tutoring during the second attempt.



absence of random assignment. We used an archival quasi-experimental approach to compare course grades among students who received tutoring only during their second attempt against appropriately selected controls who did not receive tutoring during either the first or second attempt.

The results of Phase III demonstrate that peer tutoring can be shown to have a significant *estimate causal* impact on improvements in students' grades from a first attempt of a course to a second repeated attempt. Statistical analyses in Phase III of the present study revealed that this estimated effect was strongest when adjusting for unobserved variability in the pretest measure (in this case, using grades at first attempt), indicating that tutoring likely remains effective even in situations in which students are not repeating a course. These findings are particularly interesting because students who sought tutoring began with a lower mean grade for their first attempt (a grade of 0.87 compared to 1.26 for controls), but ended with a higher mean grade for their repeated attempt (2.55 compared to 2.41 for controls). This finding fits with the work of Walker and Dancy (2007), who showed that students who use tutoring services generally tend to get lower grades than their classmates in the first place. In other words, tutoring is not only associated with improvement in students' final grades from first attempt to second attempt of a course, but these improvements significantly exceed the increase in grades for counterpart controls. This is a remarkable finding, especially given the significant limitations that must be overcome to carry out a statistically robust analysis of an active tutoring program in vivo. Results suggest that the statistically estimated effect of tutoring is closely associated with successful course outcomes.

Another important finding of Phase III is the fact that the presence of tutoring was associated with success primarily for first-generation college students. Specifically, first-generation students who chose to receive tutoring during their second attempt started out with the lowest grades after first attempt of the

course compared to all other categories of participants. However, after their second attempt, while receiving tutoring, first-generation students' grades were the highest compared to first-generation students who did not receive tutoring and continuing-generation students in either the tutoring or no-tutoring categories (see Figure 2). This finding emphasizes the estimated impact that tutoring can have on first-generation students and the importance of providing first-generation students with access to appropriate services to increase their advantages to meet their fullest potential.

As outlined by Forbus, Newbold, and Mehta (2011), first-generation college students face a myriad of unique challenges that put them at higher risk of not persisting to graduation than students who have at least one parent who completed a college degree. These challenges include a marked reduction in the amount of academic preparation experienced prior to entering college, a greater likelihood to be exposed to the demands and uncertainty of a low-socioeconomic-status background, reduced familiarity with the university environment and so-called "college knowledge," as well as significantly less social support from family members and friends (Forbus et al., 2011). Personal factors that inhibit their pathway to success include lower self-esteem and a reduced sense of self-efficacy, which can easily lead to lower ambitions for both academics and career goals (Forbus et al., 2011).

The work of Fischer (2007) and of Berger and Malaney (2001) seemed to indicate that students who become more involved in the university community, through participation in various campus organizations, are far more likely to feel attached to the institution, experience a sense of satisfaction with their collegiate involvement, and to achieve higher academic outcomes. Similarly, the results of Phase III of this study confirm that tutoring is empirically associated with successful course outcomes for first-generation college students. This finding underlines the recommendations of Engle and Tinto (2008) that tutoring is one service that institutions should prioritize in their various programmatic offerings to help first-generation students effectively transition into college life. This is especially true for institutions that primarily serve first-generation college students, although more in-depth research is called for to further investigate what makes tutoring, in particular, so effective for first-generation students.

GENERAL DISCUSSION

Results from the present study demonstrate that providing access to a structured tutoring program is closely associated with particular benefits for students' academic achievement at the undergraduate level. As seen in Phase I, students who participated in the tutoring program in this study perceived that the services meaningfully impacted many different academic objectives and, ultimately, their final course grades. As shown in Phase II, the use of a tutoring contract to encourage regular and frequent use of the service appears to be an influential strategy that is related to the effectiveness of peer tutoring programs. Phase III revealed that for students who met the requirements of the tutoring contract, the structured peer tutoring intervention was estimated to have a significant impact on course

outcomes when comparing first to second attempts of the same course. Results show that students who utilized the tutoring program during their second attempt in a course demonstrated greater academic improvements than their classmates who did not receive tutoring from the program. These findings provide a strong base of empirical support that endorses the continued implementation of highly structured academic tutoring services at the postsecondary level. We discuss the implications of these results, limitations to the findings, and suggestions for future research in detail.

The results of Phase I provide greater insight into the student experience of the effectiveness of tutoring. Overall, students responding to the survey reported that tutoring positively impacted their understanding of the course content, completion of assignments, and confidence as students, in addition to the other factors outlined in Table 1. These findings are in line with previous qualitative studies that found that students report that tutoring provides greater gains in knowledge, independent thinking, and learning passion (Chen & Liu, 2011), is an extremely helpful service (American River College, 1993), and is an enjoyable and beneficial experience (Saunders, 1992). Additionally, the finding that 88.2% of respondents indicated that tutoring had improved their grade at least to some degree is remarkably similar to the findings of Hendriksen et al. (2005) who also found that 88% of their respondents reported improved grades because of having been tutored. This finding is encouraging and also fits nicely with the results of Phase II and Phase III.

The results of Phase II support previous findings that the amount of peer tutoring is positively correlated with final course grades (Munley et al., 2010; Sobral, 2002), which speaks to the academic benefits associated with regular tutoring. Additionally, we observed positive correlations between course grades and the number of weeks that students attended tutoring, how early in the term they accessed tutoring services, overall GPA, and whether conditions of the tutoring contract were met. Interestingly, the number of weeks active in tutoring demonstrated a stronger relationship with final course grades than total hours of tutoring received. The results also show significantly higher grades for students who chose to meet the tutoring contract over students who received tutoring but did not meet the contract. These findings speak to the benefits associated with adhering to a formal contract, which might encourage students to distribute their practice more evenly throughout the academic term. As had been shown by Dunslosky et al. (2013), distributed practice is one the most highly effective study strategies available to students. The findings of Phase II of the present study highlight the importance of tutoring begun early in the term and maintained on a regular schedule throughout the academic term.

Findings from Phase III are consistent with previous literature highlighting the academic benefits of peer tutoring at the postsecondary level. Specifically, as demonstrated by Lake (1999), students who receive tutoring have significant advantages in their academic outcomes over comparison students. In the present study, these advantages of being tutored were described using an empirical estimation of the treatment effect, which is an especially meaningful contribution to the extant literature. Most notably, Phase III demonstrated that students who received tutoring during their second attempt of a course not only achieved higher

course grades than their counterpart controls, but, as shown in Figure 1, this finding represents a crossover effect because these students received lower grades during their first attempt to begin with. This is an especially meaningful finding given the inability of previous research to demonstrate or estimate the effect that tutoring can have on final course outcomes.

Interestingly, tutoring is associated with particular benefits for first-generation college students, which makes sense, at least in part, given the highly structured nature of the tutoring model evaluated in this study. Specifically, as shown in Figure 2, the results of the factorial ANCOVA in Phase III demonstrated that, although first-generation college students who sought tutoring had the lowest final grades after their first attempt, the act of receiving tutoring resulted in these same students receiving the highest grade during their second attempt compared to first-generation students who did not seek tutoring and continuing-generation students in both the tutored and not-tutored categories. This remarkable finding sheds even more light on the importance of institutions providing quality academic services to first-generation college students.

At least part of why this phenomenon is occurring could reasonably be attributed to the fact that being required to attend regular tutoring sessions with a peer tutor provides at least one meaningful campus connection for these first-generation students. Such connections, as outlined in several articles (Berger & Malaney, 2001; Fischer, 2007; Forbus et al., 2011), seem to be critical to overcoming the obstacles typically associated with first-generation college student status. This finding also seems to justify why, in 1997, tutoring was the most common service offered by TRiO Student Support Services, federal programs that primarily serve first-generation, low-income, and disabled college students (Muraskin, 1997). Additionally, the findings of Phase I showed that students perceived tutoring to have positively impacted their academic confidence. This type of boost in confidence could also be one of the many factors that causes tutoring to be so beneficial to first-generation students, given that academic self-esteem and self-efficacy have been identified as particularly problematic factors for these students (Forbus et al., 2011). Because, as Ryan, Pintrich, and Midgley (2001) point out, it is generally unclear in the literature whether students who need the most help are actually seeking it, these findings emphasize the importance of institutions going out of their way to ensure that first-generation college students are well informed about the availability and benefits of tutorial services and that they know how to access those services effectively.

Limitations

The most important finding within the present study was demonstrated using a nonequivalent comparison group design, which allowed for the empirical estimation of a *causal* relationship between receiving tutoring and improving course grades. Given that most research related to tutoring involves the use of operational tutoring programs that necessarily allow students to self-select participation in the service or not, the nonequivalent comparison group design is perhaps one of the best ways to assess the effectiveness of a tutoring program in vivo, especially when

compared to qualitative and correlational methods. However, it is still important to clarify that this does not entirely overcome the limitation of self-selection. The design seems to have desirable advantages when random assignment is not possible and might deserve greater attention in future studies, but it will ultimately continue to be limited by the self-selected nature of participating in tutoring.

The archival nature of the methodology utilized in Phase III unfortunately makes it impossible to account for the vast array of other academic success strategies and resources that might have been employed and accessed by participants in both the experimental and control groups during the academic terms of the study. For example, students in either group could have attended departmental tutoring labs, peer-facilitated study groups, personally funded tutoring sessions, or might have accessed resources such as study guides, online tutoring videos, or individual appointments with the faculty member. Future studies could benefit from greater control over such variables.

A further limitation of both the current study and those described in previous literature is that it is difficult, if not impossible, to quantitatively control for the professional quality of program personnel and training. Although certain training models are commonly used (e.g., CRLA), many of these programs allow for individual program coordinators to adapt their trainings in a manner that suits their individual strengths and program needs. Although this model of flexibility is an asset to training models like CRLA, the lack of standardized training and evaluation methods for peer tutors likely accounts for a great deal of the variability observed in tutoring outcomes between various programs. Similarly, although programs (such as the one evaluated in the present study) provide monetary compensation for tutors, there are other programs that utilize experienced faculty members or choose to rely solely on volunteer tutors (Carmody & Wood, 2009), creating even more variation in the literature related to tutoring. Future research is needed to examine these potential differences and their effect on student outcomes.

Implications and Conclusion

Because recent arguments have called into question the sustainability of personal tutoring (Myers, 2008), continued contributions to the extant literature regarding the effectiveness of tutoring are in critical demand. As demonstrated by the Delta Cost Project (Wellman et al., 2009) and Corash and Baker (2009), providing learning assistance such as tutoring was not only related to higher rates of student persistence, but that these higher rates of persistence generated enough revenue to offset the costs of providing the services in the first place (see also Lee, 1988). Given the results of the present study, the continued funding of undergraduate tutoring programs not only seems warranted but is a highly desirable strategy for institutions that seek to impact student success, especially those that primarily serve first-generation college students. Future research that investigates similar issues can meaningfully contribute to an academic culture in which tutoring services are seen as a fundamental and necessary offering by any institution dedicated to ensuring that students persist and ultimately graduate.

ABOUT THE AUTHORS

Mitchell Colver received his BA in Psychology/Music from BYU-Hawaii, his MS in Experimental Psychology from Eastern Washington University, and is currently pursuing a doctorate in Curriculum and Instruction from Utah State University. He has worked in the field of education for over eight years, primarily within collegiate level student services.

Trevor Fry is a native of Spokane, Washington. He received his BA in Addiction Studies and his MS in Experimental Psychology from Eastern Washington University and is currently pursuing a doctorate in Organizational Psychology from the Florida Institute of Technology.

REFERENCES

American River College. (1993). *A.R.C. Beacon project: Student catalyst program—peer assisted learning; first semester summary report*. Sacramento, CA: American River College.

Annis, L. F. (1983). The processes and effects of peer tutoring. *Human Learning: Journal of Practical Research & Applications*, *2*, 39–47.

Arco-Tirado, J. L., Fernández-Martín, F. D., & Fernández-Balboa, J.-M. (2011). The impact of a peer-tutoring program on quality standards in higher education. *Higher Education*, *62*(6), 773–788. doi:10.1007/s10734-011-9419-x

Berger, J., & Malaney, G. (2001, March). *Assessing the transition of transfer students from community colleges to a university*. Paper presented at the Annual Conference of the National Association of School Psychologists, Seattle, WA.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York,NY: David McKay.

Bryer, J. (2012). Peer tutoring program for academic success of returning nursing students. *The Journal of the New York State Nurses' Association*, *43*(1), 20–22.

Carmody, G., & Wood, L. (2009). Peer tutoring in mathematics for university students. *Mathematics & Computer Education*, *43*(1), 18–28.

Cepeda, N., Pashler, H., Vul, E., Wixted, J., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354–380. doi:10.1037/0033-2909.132.3.354

Chaney, B. W. (2010). *National evaluation of student support services: Examination of student outcomes after six years. Final report*. Washington, DC: Office of Planning, Evaluation and Policy Development, US Department of Education.

Chen, C., & Liu, C. C. (2011). A case study of peer tutoring program in higher education. *Research in Higher Education Journal*, *11*, 1–10.

Cohen, P. A., Kulik, J. A., & Kulik, C. L. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, *19*(2), 237–248. doi:10.3102/00028312019002237

College Reading and Learning Association. (2015). *International Tutor Training Program Certification (ITTPC)*. Retrieved from http://crla.net/index.php/certifications/ittpc-international-tutor-training-program

Comfort, P. (2011). The effect of peer tutoring on academic achievement during practical assessments in applied sports science students. *Innovations in Education and Teaching International*, *48*(2), 207–211. doi:10.1080/14703297.2011.564015

Cooper, E. (2010). Tutoring center effectiveness: The effect of drop-in tutoring. *Journal of College Reading and Learning*, *40*(2), 21–34. doi:10.1080/10790195.2010.10850328

Corash, K., & Baker, E. D. (2009). *Calculating the productivity of innovation: Using a simplified cost-benefit analysis to promote effective practice*. Denver, CO: Colorado Community College System.

Dioso-Henson, L. (2012). The effect of reciprocal peer tutoring and non-reciprocal peer tutoring on the performance of students in college physics. *Research in Education*, *87*(1), 34–49. doi:10.7227/RIE.87.1.3

Dugard, P., & Todman, J. (1995). Analysis of pre-test post-test control group designs in educational research. *Educational Psychology*, *15*(2), 181–198. doi:10.1080/0144341950150207

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4–58. doi:10.1177/1529100612453266

Engle, J., & Tinto, V. (2008). *Moving beyond access: College success for low-income, first-generation students*. Washington, DC: The Pell Institute.

Farmer, J., Lachter, G. D., Blaustein, J. J., & Cole, B. K. (1972). The role of proctoring in personalized instruction. *Journal of Applied Behavior Analysis*, *5*(4), 401–404. doi:10.1901/jaba.1972.5-401

Fischer, M. (2007). Settling into campus life: Differences by race/ethnicity in college involvement and outcomes. *The Journal of Higher Education*, *78*(2), 125–156. doi:10.1353/jhe.2007.0009

Forbus, P., Newbold, J. J., & Mehta, S. S. (2011). A study of non-traditional and traditional students in terms of their time management behaviors, stress factors, and coping strategies. *Academy of Educational Leadership Journal*, *15*, 109–125.

Fuller, J. L., & Denehy, G. E. (1972). Student tutoring in preclinical technique laboratory. *Journal of Dental Education*, *36*(9), 18–24.

Hendriksen, S. I., Yang, L., Love, B., & Hall, M. C. (2005). Assessing academic support: The effects of tutoring on student learning outcomes. *Journal of College Reading & Learning*, *35*(2), 56–65. doi:10.1080/10790195.2005.10850173

Higgins, B. (2004). Relationship between retention and peer tutoring for at-risk students. *The Journal of Nursing Education*, *43*(7), 319–321.

Holliday, T. (2012). Evaluating the effectiveness of tutoring: An easier way. *Learning Assistance Review (TLAR)*, *17*(2), 21–32.

House, J. D., & Wohlt, V. (1990). The effect of tutoring program participation on the performance of academically underprepared college freshmen. *Journal of College Student Development*, *31*(4), 365–370.

Kimmel, A. (1988). *Ethics and values in applied social research* (Vol. 12). New York: Sage.

Lake, D. A. (1999). Peer tutoring improves student performance in an advanced physiology course. *American Journal of Physiology*, *21*(1), 86–92.

Lee, R. E. (1988). Assessing retention program holding power effectiveness across smaller community colleges. *Journal of College Student Development*, *29*(3), 255–262.

Lidren, D. M., Meier, S. E., & Brigham, T. A. (1991). The effects of minimal and maximal peer tutoring systems on the academic performance of college. *Psychological Record*, *41*(1), 69.

May, H.. (2012). Nonequivalent comparison group designs. In H. Cooper Ed., *APA handbook of research methods in psychology: Vol 2. Research designs* (489–509). Washington, D.C.: American Psychological Association. doi:10.1037/13620-026

Metcalfe, R. (1992). The assessment of proctoring. *Aspects of Educational and Training Technology Series*, *25*, 42–46.

Munley, V. G., Garvey, E., & McConnell, M. J. (2010). The effectiveness of peer tutoring on student achievement at the university level. *The American Economic Review*, *100*(2), 277–282. doi:10.1257/aer.100.2.277

Muraskin, L. (1997). *"Best practices" in student support services: A study of five exemplary sites*. Washington, DC: U.S. Department of Education. Retrieved from http://files.eric.ed.gov/fulltext/ED411739.pdf

Myers, J. (2008). Is personal tutoring sustainable? Comparing the trajectory of the personal tutor with that of the residential warden. *Teaching in Higher Education*, *13*(5), 607–611. doi:10.1080/13562510802334988

Mynard, J., & Almarzouqi, I. (2006). Investigating peer tutoring. *Elt Journal*, *60*(1), 13–22. doi:10.1093/elt/cci077

National Conference of State Legislatures. (2014). *Performance based funding for higher education*. Retrieved from http://www.ncsl.org/research/education/performance-funding.aspx

Reinheimer, D., & McKenzie, K. (2011). The impact of tutoring on the academic success of undeclared students. *Journal of College Reading & Learning*, *41*(2), 22–36. doi:10.1080/10790195.2011.10850340

Ryan, A. M., Pintrich, P. R., & Midgley, C. (2001). Avoiding seeking help in the classroom: Who and why? *Educational Psychology Review*, *13*(2), 93–114. doi:10.1023/A:1009013420053

Santee, J., & Garavalia, L. (2006). Peer tutoring programs in health professions schools. *American Journal of Pharmaceutical Education*, *70*(3), 70–10. doi:10.5688/aj700370

Saunders, D. (1992). Peer tutoring in higher education. *Studies in Higher Education*, *17*(2), 211–218. doi:10.1080/03075079212331382677

Sobral, D. T. (2002). Cross-year peer tutoring experience in a medical school: Conditions and outcomes for student tutors. *Medical Education*, *36*(11), 1064–1070. doi:10.1046/j.1365-2923.2002.01308.x

Topping, K. J. (1996). The effectiveness of peer tutoring in further and higher education: A typology and review of the literature. *Higher Education: The International Journal of Higher Education and Educational Planning*, *32*(3), 321–345. doi:10.1007/BF00138870

Topping, K. J., Watson, G. A., Jarvis, R. J., & Hill, S. (1996). Same-year paired peer tutoring with first year undergraduates. *Teaching in Higher Education*, *1*(3), 341–356. doi:10.1080/1356251960010305

Trevino, F. M., & Eiland, D. C., Jr (1980). Evaluation of a basic science, peer tutorial program for first-and second-year medical students. *Academic Medicine*, *55*(11), 952–953. doi:10.1097/00001888-198011000-00009

Walker, K. N., & Dancy, M. H. (2007). Investigation and evaluation of a physics tutorial center. *AIP Conference Proceedings*, *883*, 137–140. doi:10.1097/00001888-198404000-00004

Walker-Bartnick, L. A., Berger, J. H., & Kappelman, M. M. (1984). A model for peer tutoring in the medical school setting. *Academic Medicine*, *59*(4), 309–315. doi:10.1097/00001888-198404000-00004

Wellman, J. V., Desrochers, D. M., Lenihan, C. M., Kirshstein, R. J., Hurlburt, S., & Honegger, S. (2009). *Trends in college spending: Where does the money come from? Where does it go?* Washington, DC: Delta Cost Project at American Institutes for Research.