Teaching Mathematics and its Applications: An International Journal of the IMA (2023) **00**, 1–24
https://doi.org/10.1093/teamat/hrac026 Advance Access publication

Towards research-based organizational structures in mathematics tutoring centres

Cameron Byerley[†], Carolyn Johns[‡,*], Deborah Moore-Russo[§], Brian Rickard[‖], Carolyn James[¶], Melissa Mills[**], Behailu Mammo[††], Janet Oien[‡‡], Linda Burks[§§], William Heasom[‖‖], Melissa Ferreira[‖‖], Cynthia Farthing[¶¶] and Daniel Moritz[***]

[†]*University of Georgia, Athens, Georgia, 30602, USA*
[‡]*The Ohio State University, Columbus, Ohio, 43210, USA*
[§]*University of Oklahoma, Norman, Oklahoma, 73019, USA*
[‖]*University of Arkansas, Fayetteville, Arkansas, 72701, USA*
[¶]*University of Portland, Portland, Oregon, 97203, USA*
[**]*Oklahoma State University, Stillwater, Oklahoma, 74078, USA*
[††]*Hofstra University, Hempstead, New York, 11549, USA*
[‡‡]*Colorado State University, Fort Collins, Colorado, 80523, USA*
[§§]*Santa Clara University, Santa Clara, California, 95053, USA*
[‖‖]*Villanova University, Villanova, Pennsylvania, 19085 USA*
[¶¶]*University of Iowa, Iowa City, Iowa, 52242, USA*
[***]*University of Colorado Boulder, Boulder, Colorado, 80309, USA*

[*]*Corresponding author. Email:* johns.125@math.osu.edu

[Received June 2021; accepted December 2022]

Undergraduate mathematics tutoring centres are prevalent in many countries; however, there is limited research-based evidence on effective organizational structures for these centres. In this study, we consider two research questions. First, how can the quantitative and qualitative data from 10 mathematics tutoring centres be organized for research purposes? Second, what hypotheses do expert mathematics tutoring centre leaders generate about characteristics of effective centres given data from a sample of ten centres? We collected quantitative data from over 26,000 students taking mathematics courses at ten institutions. Data collected included college entrance exam scores, high school grade point average, number of student visits to the centre per eligible student and course letter grade. We used exploratory data analysis to look for relationships between visits to the tutoring centre, student grades and other variables. Qualitative centre characteristics that were considered include: specialist–generalist tutoring system, tutoring capacity, physical layout, relationships between tutors and mathematics instructors and extent of tutor training. We used the Delphi process to generate testable

hypotheses from the data, such as the following: (1) The more courses a tutor is responsible for tutoring the more likely it is that the tutor will struggle to answer student questions, when the difficulty level of the courses is roughly the same. (2) Centres with more specialized tutor models have more visits per student than centres with generalized tutor models. The preceding two hypotheses, along with the other generated hypotheses, have been identified by the experts participating in this study as plausible based on professional experience, exploratory data analysis and inferences based on prior research on tutoring. This study has not rigorously shown the validity of these hypotheses; rather it lays the groundwork for future investigations to determine what combination of features characterize an effective tutoring centre.

1. Introduction

A mathematics tutoring centre (often referred to as a tutoring centre or just centre when mathematics tutoring is clear from context) is ‘a place on a university campus where students enrolled in a mathematics course can get optional out-of-class resources to support their learning’ (Mills *et al.*, 2020, p.3). Mathematics tutoring centres, also called mathematics support centres by some, are extremely common in the Republic of Ireland, the UK, the USA and Australia (MacGillivray & Croft, 2009; Johnson & Hanson, 2015; Cronin *et al.*, 2016; Grove *et al.*, 2020). We define tutors as individuals who provide unstructured out-of-classroom help to students typically in a centre. Despite the prevalence of tutoring centres in many countries and a large body of literature on characteristics of effective tutoring practices (Schoenfeld *et al.*, 1992; Graesser & Person, 1994; Chi, 1996; Lepper & Woolverton, 2002; Ryals *et al.*, 2019), there is limited research-based evidence on effective tutoring centre organizational structures (Matthews *et al.*, 2013; Mills *et al.*, 2020). Mills *et al.* (2020) noted that tutoring centres in the USA ‘vary a great deal in their practices, resources and organizational structures’ (p. 1). Increasingly, calls have been made to further examine the effectiveness of tutoring centres in order to understand centre best practices (Moore-Russo *et al.*, 2018; Lawson *et al.*, 2020; Mills *et al.*, 2020).

Much of the existing research on tutoring centre effectiveness has focused on metrics such as number of visits (Berkopes & Abshire, 2016) and does not give research-based insight into what made the centre effective. There are a few studies that offer evidence that tutoring centre visits have a positive effect on student success while controlling for other variables (Xu *et al.*, 2001; Berry *et al.*, 2015; Byerley *et al.*, 2018; Rickard & Mills, 2018; Rylands & Shearman, 2018; Jacob & Ní Fhloinn, 2019; Mullen *et al.*, 2021). Since each of these studies was conducted at single institutions, it is difficult to know what contributed to their effectiveness. We are interested in knowing more than if (and to what degree) a tutoring centre is effective; we are also interested in the relationship between a centre’s organizational structure and its effectiveness.

Our investigation of the relationship between organizational structure and success took place over a number of years during mathematics tutoring centre leader conferences and working group meetings. In the first year of work, a few centre leaders worked together to collect and analyze quantitative data under the leadership of the team’s statistician. Initial reports of quantitative work were published and used as templates for other centres’ data collection and analysis (Byerley *et al.*, 2018; Rickard & Mills, 2018). In the next year, six of the authors of this paper focused on identifying and defining structural features of centres to lay a foundation for investigation of characteristics of effective centres (Byerley *et al.*, 2019). Then, using both the shared definitions of organizational structures and shared methods of quantitative analysis, we (the authors of this paper) collected and organized qualitative and quantitative data describing ten mathematics tutoring centres (Byerley *et al.*, 2020). Finally, we analyzed the qualitative and quantitative data from ten centres to generate testable hypotheses about characteristics of effective centres using the Delphi process (Pill, 1971). This paper synthesizes the data and provides hypotheses about characteristics of effective mathematics tutoring centres.

We are particularly excited about this work because there is untapped potential to improve tutoring centre organizational structures based on research. Many of our proposed structures do not take substantial amounts of time or money to implement compared to other educational interventions, such as changing classroom instruction or university policies. In our experience, undergraduate and graduate tutors are usually willing to follow directions and implement new ideas because they are so new to the practice of teaching and tutoring; hence, they are willing to accept advice (Johns, 2020). Because not much is known about effective centre structures and because it is relatively inexpensive and quick to change the structures in our experience, we believe that the line of research proposed in this paper offers substantial benefits for students relative to the investment (Mills *et al.*, 2017).

2. Theoretical framework and literature review

A subset of the authors of this paper identified and defined six dimensions that differed among our tutoring centres that serve as a framework in this paper (Byerley *et al.*, 2019). To determine the organizational identity that is associated with mathematics centres we looked at the central, stable features of ten mathematics centres that make them distinctive (Gioia *et al.*, 2013). We note that because our centres exist within the USA, the USA context shapes the operations of our tutoring centres that could differ from the operations of tutoring centres elsewhere in the world, which are shaped by their own national contexts. All six authors of Byerley *et al.* (2019) were well positioned to describe dimensions of organizational identity because they were actively involved in their universities' mathematics tutoring centres, attended a national conference for tutor centre leaders, participated in weekly or monthly online meetings with other tutor centre directors and led or attended tutoring centre working groups. Our understanding of tutoring centre structures is built on our frequent interaction with our universities' tutoring centres, notes from conferences and online meetings and a shared digital resource library. The dimensions we identified are: (1) specialist–generalist tutoring spectrum, (2) strength of relationship between tutoring centre and mathematics instructors, (3) type and extent of tutor training, (4) types of tutoring services, (5) physical layout and location and (6) tutoring capacity. These dimensions were used to organize the literature review, the collection of qualitative and quantitative data and the creation of figures used in the Delphi process. Although we collected data on each dimension and considered each dimension when writing hypotheses, the top hypotheses that are the foci of the paper are related to dimensions 1, 2 and 3. In the Results section, we spend more time reviewing the literature related to the most popular hypotheses that emerged from the Delphi process.

2.1 Specialist–generalist tutoring spectrum

In a specialist tutor model, a mathematics tutor is assigned to tutor one course or a small number of related courses. A specialized tutor becomes familiar with the homework problems, student mistakes, homework solutions, the syllabus and expectations for testing. Some specialized tutors also serve as learning assistants in the course that they tutor in the centre. This means that they regularly attend lectures to assist the classroom instructor (see Goertzen *et al.* (2011) for further definition of learning assistants). We consider tutors who are also instructors, graders or teaching assistants for a course highly specialized because of the extensive knowledge of a specific course. Here, we define instructors as individuals who are the instructor of record for the course, that is, they give lectures and are responsible for assigning grades. We define graders as individuals who are not instructors but perform grading duties for a course. Finally, we define teaching assistants as individuals who formally meet with a subset of the course students, but are not the instructor of record. Some tutors specialize in a small number of related courses, but those tutors do not necessarily attend the courses or interact with the instructors frequently. For example, specialized tutors might tutor the first two courses in a calculus sequence but not other courses.

A generalist mathematics tutor typically tutors many courses. They answer student questions in the order the questions were asked and thus shift between answering questions for different courses multiple times per hour.

Centres can also have a mix of specialist and generalist tutors. For example, at some centres, undergraduates tutor all courses the centre offers while graduate tutors hold office hours for the class for which they are instructors of record. There are many organizational structures that contribute to a tutor's development of in-depth knowledge of a course; so, it is non-trivial to place a tutoring centre on a spectrum from specialized models to generalized models. Although studying specialization and generalization in tutoring centres is a relatively new area of interest, specialization and generalization has been studied in organizational structures for businesses for many years (Weisbord, 1976). For example, Weisbord (1976) observed, 'in-depth competence erodes rapidly in each specialty, for generalists cannot keep up with everything' (p. 438).

2.2 *Tutor training*

In both Europe and the USA, there is variation in the amount and content of training provided to tutors (Cronin *et al.*, 2016; Mills *et al.*, 2020) with training typically lasting from 1 to 10 h. Lawson *et al.* (2020) noted that there is recognition of the importance of training in the UK and Ireland with free training materials being developed (Croft & Grove, 2016).

Proponents of tutor training argue tutors serve multiple roles, which require them to be knowledgeable in a wide range of areas (Ireland, 2006; Croft & Grove, 2016; Delderfield & McHattie, 2018; Lawson *et al.*, 2020), and mathematical knowledge alone is not sufficient (Gillard *et al.*, 2011; Walsh, 2017). Tutors also should understand student thinking, work with a diverse range of students, address students' emotional needs, teach study strategies and use pedagogically appropriate questioning techniques (Fitzmaurice *et al.*, 2016; Burks & James, 2019). Centre leaders often have the autonomy to choose the topics that they cover in tutor training, and their choices can vary based on the centre leader's philosophy of tutoring, the specific needs of the tutors, budgeting issues and values of the department. There is no formal research on the impact of different types of tutor training on centre effectiveness.

2.3 *Strength of relationship between tutoring centre and mathematics instructors*

Stronger relationships between the centre and the instructors can benefit students. Cronin & Meehan (2021) examined the use of providing lecturers with a summary of student queries raised in the tutoring centre. They found that lecturers reported this was valuable formative feedback. In a study of out-of-school supports for secondary students, there were similar findings that suggest the importance of having teachers involved in the support centre (MacBeath *et al.*, 2001). The authors suggested out-of-school support worked better when it was 'built-in' to the overall design of the courses rather than 'bolted-on' as a disconnected extra service.

The relationship between mathematics tutoring centre leaders, mathematics faculty and tutors varies by institution. For example, some mathematics tutoring centres are located within a larger umbrella university support centre (Starkings, 2002; Gordon, 2004) while others are situated within the mathematics department (Lawson *et al.*, 2020). Grove *et al.* (2018) found that 60% of 48 UK institutions surveyed integrated mathematics and statistics support with other support services, and only 27% were managed entirely by an academic department. Only 17 of 51 UK institutions surveyed used course instructors as mathematics centre tutors (Grove *et al.*, 2019). Mills *et al.* (2020) found 76% of 75 tutoring centres surveyed in the USA reported collaboration with mathematics faculty, and 65% reported tutoring by graduate students and faculty. When instructors spend their office hours at the centre tutoring, there is a greater potential for a relationship between instructors, undergraduate tutors and the centre director.

2.4 *Types of tutoring services*

Some centres in this study focus on a particular type of mathematics, such as calculus, and only serve a few courses, while other centres serve over 20 different courses ranging from developmental mathematics to linear algebra. Centres can offer drop-in tutoring, scheduled one-on-one tutoring or a combination of services. A potential benefit of drop-in tutoring is that students work together and build relationships with classmates. A potential benefit of having an appointment is that a student gets focused attention for a longer period of time. It is plausible that drop-in centres attract different types of students than appointment-based centres. For example, high performing students are often comfortable working with friends at a drop-in centre. The high-performing student, who only has occasional questions for a tutor, might not book an appointment for an hour of private help if the private appointments are advertised as being for struggling students.

2.5 *Physical layout and location*

The size and quality of spaces used by tutoring centres vary widely in the USA. Seating capacity varies from less than 10 seats to over 100 (Mills *et al.*, 2020). There are also variations in the tutoring centre's location on campus, how far the students typically must travel to attend the centre and how close the centre is to the mathematics department. In addition, the physical space may offer separate spaces dedicated to specific populations (Lawson *et al.*, 2020).

2.6 *Tutoring capacity*

Mills *et al.* (2020) defined tutoring capacity as 'a measure of how many tutoring hours per year are offered per eligible student for different sizes of universities' (p. 12). Note that tutoring hours per year is not the same as the number of hours a centre is open because a centre often has multiple tutors working at once. We chose to examine tutoring hours per year rather than the number of hours a centre is open to focus on the opportunities students have to interact with a tutor, rather than the opportunities they have to be in the centre. If a centre has limited hours that do not work well with students' schedules, it will likely negatively impact visitation but we did not consider that issue because all centres in this study were open many hours each week. The average tutor hours per year per eligible student varied from 4.34 h at small institutions to 1.09 h at large institutions in the Mills *et al.* (2020) study of 75 USA centres. It is unknown if centres with increased tutoring capacity are more effective because it is possible that groups of students at busy centres will work together to solve many of their own problems and learn from that process.

3. Methods

This project focuses on research questions of interest to practitioners as suggested by Cai *et al.* (2019). In May 2017, the USA National Science Foundation funded a three-day conference for mathematics tutoring centre leaders where the group brainstormed topics of research. Consistent with community-based participatory research (Brady, 2015), research questions should be generated by practitioner questions; practitioners and researchers mutually benefit from the ongoing research relationships; and the research outcomes should have direct applicability to the practitioners' work. To generate knowledge for practitioners (in this case, the mathematics tutoring centre leaders) using their experiences, we used the Delphi process to address the following research questions (Skulmoski *et al.*, 2007):

1. How can the quantitative and qualitative characteristics of ten mathematics tutoring centres be organized for research purposes?
2. What hypotheses do expert mathematics tutoring centre leaders generate about characteristics of effective centres given data from 10 centres?

We describe each major stage of the research in Sections 3.1 to 3.3.

3.1 *Collecting qualitative summaries of centres*

After creating initial definitions of six dimensions, a subset of the authors thought might be related to characteristics of successful centres (Byerley *et al.*, 2019), each author wrote a qualitative description of his or her centre with attention to each dimension. The leaders described both strengths and weaknesses of their centres and submitted information to the lead author. She blinded each description and renamed centres using animal names. The descriptions of the centres included a number of nuanced observations and ranged from five to ten pages in length. After the group read the descriptions of each centre, we identified additional details that we wanted to know about each centre. A subgroup created a survey, and each centre leader provided further information. For example, the survey specifically asked how many courses new and experienced tutors were responsible for tutoring to help us place the centres on the specialized to generalized spectrum.

3.2 *Collecting quantitative summaries of tutoring centres*

The methods used for quantitative evaluation of tutoring centres' impact on student success have grown in sophistication over the last 20 years as researchers attempt to control for self-selection bias. As it is not possible to randomly assign students to attend or not attend a tutoring centre, researchers must consider the possibility of self-selection bias. For example, if more motivated students are more likely to attend, relationships between centre use and higher course grades might be due to student motivation, not the help received at the centre. MacGillivray & Croft (2011) advocated for the use of general linear regression to evaluate centres. They wrote 'the essential concept is to compare performance relative to a base measure for those who used [the tutoring centre] with the same relative performance for those who did not' (p. 200). They suggested use of the students' prior grade point average (GPA), results on a first assessment and diagnostic test data as possible baseline measures. Although MacGillivray & Croft (2011) noted that general linear models are useful for analyzing the relationship between many variables and student performance, they only noted one study of tutoring centres (MacGillivray & Croft, 2009) that used general linear models. Matthews *et al.* (2013) conducted a literature review on the evaluation of tutoring centres, and most of the studies reviewed did not control for self-selection bias. Since 2009 more research groups have use multiple linear regression to evaluate tutoring centres (Berry *et al.*, 2015; Byerley *et al.*, 2018; Rickard & Mills, 2018; Rylands & Shearman, 2018; Jacob & Ní Fhloinn, 2019). Three authors, who are centre leaders, previously published a quantitative analysis of their centres' data that found positive relationships between centre attendance and course grades after controlling for other variables (Byerley *et al.*, 2018; Rickard & Mills, 2018). Since relatively few institutions have published a linear regression analysis of their centres it is not possible to generalize the result from studies of one institution to new contexts. For example, Walker and Dancy (2006) found that students who attended a physics tutoring centre had 20% lower mean exam scores than those who never attended (p. 138). They hypothesized that students who struggled self-selected to use the tutoring centre.

We collected quantitative data from over 26,000 students who were enrolled in a mathematics course at the ten institutions during one fall semester. One strength of our quantitative data collection is that the number of institutions involved allows us to investigate if a relationship between variables at one institution is apparent in another and then to consider how differences in centre structures impact measures of success. In order to standardize analyses across universities, participating centre leaders were surveyed to determine the data available to them. The factors that essentially all contributors reported being able to procure data for included: (a) college entrance exam scores, (b) high school GPA, (c) number of student visits to the centre per eligible student and (d) course letter grade converted to grade points. Multiple regression analyses were conducted on the data from each centre using course grades as the dependent variable and college entrance scores, high school GPA and number of student visits to the centre as the independent variables. We acknowledge that these factors represent only a small portion of factors that might influence student grades in a course. However, as these were the only data available to all centres, they represent the largest possible subset of common factors. We also acknowledge that course grades are only one measure of success, and this variable has limitations. For example, course grades do not necessarily measure the development of productive mathematical understanding. For example, in the USA, the majority of calculus tests assess procedural knowledge despite a focus on the development of conceptual understanding of calculus in educational research (Tallman *et al.*, 2016). Tutoring centres also hope to contribute to goals beyond the development of mathematical knowledge. For example, we value the retention of marginalized students in STEM and supporting students' identity growth as mathematical thinkers (Gutiérrez, 2013; Ellis *et al.*, 2016). Even though we understand that there is more to success than course grades, we know that students, instructors and departments want students to pass courses; so, we use this variable as one measure of success.

The fourth author had a PhD in statistics, and he assisted all centre leaders in analyzing their data and checking that their data sets met the assumptions for multiple linear regression (Cohen *et al.*, 2013). For example, we looked at scatterplots of the data to see if there was a roughly linear relationship between visits to the centre and course grade. There is multi-collinearity between the control variables college entrance exam scores and high school GPA, but our statistician deemed multiple linear regression was still a useful model for the data sets. Also, many other studies use both high school GPA and exam scores as predictor variables (Cohn *et al.*, 2004).

We also measured the effectiveness of a centre by computing the mean visits per eligible student and the mean visits of students who attended at least once (Matthews *et al.*, 2013). A high percent of students using a centre likely means it is well advertised and well recommended by students and faculty. A high number of return visits suggests students were satisfied with their initial experience at the centre. Visit metrics must be interpreted in light of each centre's context. For example, Cat's Precalculus and Gorilla's College Algebra courses required the students to visit the centre. Since none of the Bird, Dolphin, Goat and Fish faculty held office hours in the centre (as was the case in the other six centres), none of the office hours visits were counted to centre attendance for Bird, Dolphin, Goat or Fish.

We had to make a number of decisions about which data to include and exclude in our calculations to ensure standardization across institution. For many centres, students from any course are allowed to use the centre, including courses for which the centre may not have assigned tutors. Data were only be analyzed for students enrolled in courses for which the centre specifically provides tutors. Data for all students in a course served by the centre were collected, including those who did not visit the centre. Students with missing data and students who withdrew from the course were removed from the multiple linear regression analyses; these counts are shared in the data tables in the results. Students enrolled in multiple mathematics courses were treated as separate data points, with the number of visits to the tutoring centre split equally between the courses taken. Due to many smaller enrolment courses, analyses combined all courses within each university.

3.3 Delphi process of hypotheses generation

The ultimate goal of this research is to identify organizational structures, which support successful tutoring centres. However, tutoring centres are complex organizations and a centre leader may have difficulty isolating the impact of any single decision. Moreover, each centre leader tended to only know about his or her own centre and had difficulty generalizing beyond individual experience. Thus, it is productive to use the Delphi process to settle on viable hypotheses to test *before* designing a study specifically to test a hypothesis. In our process, each expert used the blinded qualitative and quantitative data, their experience and research literature to write hypotheses and sent the hypotheses to the lead author who blinded them for the group. The hypotheses were refined by experts in an iterative cycle of anonymous writing, editing and voting, described in detail below.

The Delphi process is a multi-stage process in which experts anonymously provide judgements that are systematically revisited until patterns of agreement emerge (McKenna, 1994). The Delphi process provides a mechanism for ‘soliciting and receiving honest expert opinions on a topic without fear of responses being impacted by unequal power dynamics, in-person group think, difference in social identities and values or history with one another’ (Brady, 2015). The use of blinded data and anonymous communication (as opposed to round-table discussion) allow those involved to focus on the task at hand and ‘substantially reduces the social-emotional behaviour often found when using other methods’ (Clayton, 1997). A Delphi process should leverage the experience of experts and move toward a shared judgement or opinion.

We used the Delphi process because it ‘can be applied to problems that do not lend themselves to precise analytical techniques but rather could benefit from the subjective judgments of individuals on a collective basis’ (Skulmoski *et al.*, 2007). The Delphi process is well suited to define issues and concepts, determine priorities and identify best practices (Fletcher & Marchildon, 2014). We had to make subjective judgements to coordinate evidence consisting of qualitative descriptions, quantitative data, research literature and professional experience. The aspects of our study that align with the Delphi process include use of: an intentional sample of experts; an emergent study design; structured, anonymous data that were collected and analyzed in iterations; and communication structures that preserved anonymity (Linstone & Turoff, 1975). We collected and used quantitative information but also adopted a qualitative approach as outlined by Fletcher & Marchildon (2014) in what they called a ‘modified and open-ended Delphi method.’ The Delphi process has been used in other mathematics education contexts to leverage expert opinions on complex topics (Manizade & Mason, 2011; Muñiz-Rodríguez *et al.*, 2017).

Following data collection, the hypothesis generation process included five primary stages shown in Fig. 1.

Stage 1: hypotheses generation. We generated the hypotheses using our professional experience, qualitative and quantitative data and knowledge of tutoring literature. In addition to writing hypotheses, we anonymously justified our hypotheses using evidence from the data, personal experience and research literature.

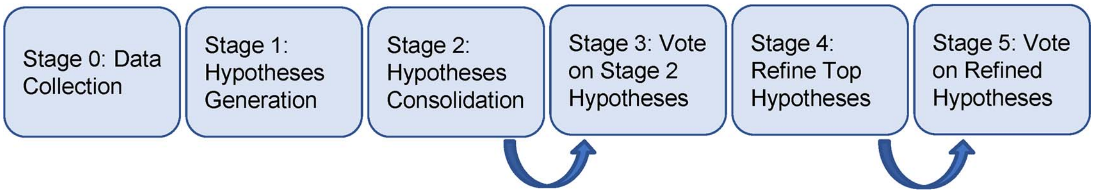

FIG. 1. The five stages of our Delphi process.

Stage 2: hypotheses consolidation. Once all centre leaders had generated hypotheses and evidence, the first and third authors organized them based on theme. The thematic approach was primarily deductive, relying on the structures of tutoring centres outlined in Byerley *et al.* (2019) for an initial framework. The result was seven themes with 19 hypotheses related to: tutor training (6), tutor–student ratios (3), social aspects of centre/tutor and student relationships (3), alternative methods to measure success (3), centre's relationship with mathematics department (2), centre attendance (1), physical location and space of centre (1).

Stage 3: vote on Stage 2 hypotheses. Eight of the 13 centre leaders[1] voted on the hypotheses they believed to be best. The 'best' hypotheses were those that were both well supported by evidence and subjective opinions based on professional experience and had strong explanatory power. The refined list of hypotheses and evidence were provided to each centre leader with the following instructions:

Highlight all hypotheses that you think are the best. Best hypotheses are supported by a combination of qualitative and quantitative data, the literature, and your expertise as a centre director. You can choose how many hypotheses to highlight.

- Indicate if you think any hypotheses are inaccurate or unsupported by data or research.
- Add your justification, counterevidence or supporting literature to each hypothesis.

The voting resulted in the following ranked hypotheses, none of which were opposed.

1. A specialized tutor model increases tutor quality, increases the number of student visits to the centre and increases student success. (seven votes)
2. Tutor training strengthens a centre and increases student success. (six votes)
3. Faculty and graduate teaching assistants holding office hours in the centre improves the effectiveness of the centre. (six votes)
4. Providing adequate space, ambiance and location for a centre leads to increased attendance and better tutoring effectiveness. (six votes, one with qualification that this was true given good tutoring occurred at centre)
5. The number of full-time employees (faculty or staff) who run the centre will have an impact on the quality of reporting and tutor training. (four votes)

Stage 4: refine top hypotheses. After extensive group discussion of the hypotheses, we realized that the wording of many of the hypotheses needed to be improved. We also realized that in our effort to combine multiple hypotheses from various members we had made each hypothesis too complex to vote on. For

[1] Some of the ten centres had more than one leader participating in this paper. Not all authors participated in each stage of the Delphi process. When not all authors participated in a stage of the Delphi process, the number of authors who participated is indicated.

TABLE 1. *Summary of hypotheses and the rank of each hypothesis. Blanks occur where no hypotheses were generated by Delphi process*

	Harder to tutor each course	More visits per student	More return visits	Larger effects of visits on grades
Tutoring more courses	1 [34]			
More specialized model		2 [33]	6 [12]	8 [9]
More tutor training			3 [25]	7 [10]
More office hours at centre		5 [13]		4 [16]

The number in brackets is the score from voting

example, the hypotheses ‘A specialized tutor model increases tutor quality, increases student visits to centre and increases student success’ made claims about three impacts of a specialized tutor model, and some centre leaders agreed with some of the claims but not all three. We revised the top three hypotheses from Stage 3 by separating them into eight more focused and testable hypotheses as described in the final voting of Stage 5.

Stage 5: vote on refined hypotheses. In the first vote, the centre leaders evaluated the hypotheses based on all of the qualitative and quantitative data. In Stage 5, they were asked to focus on the charts and tables the group designed to relate to each hypothesis. The charts and tables included both qualitative and quantitative data related to dimensions of interest. These charts answered the first research question by providing one way to organize information available to tutoring centre leaders for research. We used a forced-choice four-point Likert scale for voting to prevent centre leaders from choosing the ‘neutral’ option.

Twelve of the 13 centre leaders voted by emailing their votes to the first author. For each hypothesis, each centre leader voted twice, once based on our data and the other based on professional judgement. For each vote, *strongly agree* votes were assigned 2 points; *agree,* 1 point, *disagree,* −1 point; and *strongly disagree*, −2 points. Hypotheses earned or lost points based on each centre leader’s opinion of our data as well as based on the centre leader’s experience. The highest number of points a hypothesis could earn was 48 points. This would occur if all twelve leaders cast *strongly agree* votes that the data *and* their experience supported a hypothesis. Zero points would indicate the group equally agreed and disagreed with a hypothesis. Negative points would indicate that more centre leaders disagreed than agreed with the hypotheses.

Creating charts related to specialist and generalist hypotheses required ranking the tutoring structure. We organized the centres on a spectrum from least to most specialized using the data in Supplemental Tables 1 and 2 and provided qualitative descriptions of each centre to determine which centres structures gave the most opportunity for tutors to gain specialized course-specific knowledge. The third author initially ordered the 10 centres from most generalized to most specialized, and the first author provided suggestions and feedback until agreement between the two was reached. The first and third authors presented their list at a research meeting to the other authors, and the group accepted the order in Fig. 3. We considered a number of factors when ranking centres from most to least specialized. For example, Cat tutors were responsible for multiple versions of similar courses (i.e., tutoring both three-credit and four-credit differential calculus courses) so even if a tutor was in charge of multiple courses, they involved similar content. At Whale the tutors were expected to be willing to help with any course. However, many tutors were graduate teaching assistants who held their office hours at the centre, and the schedule listed the courses each tutor taught. Thus, students visiting the Whale centre knew which tutors had

TABLE 2. *We used linear regression to predict mathematics course letter grade point with number of visits, high school GPA and SAT or ACT*

School	Number of students	R^2	Predicted increase in grade per 1 visit	Increase in grade point per 1 grade point HS GPA	Increase in grade per 1 standard deviation SAT/ACT	Number of withdraws/ incompletes
Bird	1096	0.17	0.003	1.00***	0.26***	40
Cat	3270	0.26	0.019***	1.09***	0.59***	540
Dog	1004	0.25	0.035***	0.77***	0.49***	105
Dolphin	1070	0.15	−0.034**	1.63***	NA	87
Goat	443	0.19	−0.057***	0.57***	0.36***	18
Gorilla‡	2737	0.19	0.015***	0.67***	0.24***	447
Fish	6609	0.09	0.022***	0.71***	0.13***	639
Hamster	5151	0.17	−0.002	1.08***	0.13***	850
Horse	1971	0.12	−0.006	0.20***	0.38***	69
Whale	3453	0.23	0.016***	0.86***	0.26***	360

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$; ‡Gorilla U used HS GPA in mathematics courses rather than overall HS GPA.

specialized knowledge about a course. The students at Whale spent a lot of time receiving tutoring from someone who was teaching the course they were taking. We also had to decide how to rank centres that had both specialized and generalized tutors. Undergraduate tutors at Gorilla tutored eight courses and graduate tutors held their office hours at the centre and tutored only the course they taught. To decide how specialized Gorilla was compared to the other centres, we determined what proportion of the tutoring was done by graduate students. Although we feel confident our list is reasonable for the purpose of hypotheses generation, we acknowledge that future research investigating specialist or generalist models should take care to collect data to see if students are receiving help from specialized instructors or a generalist tutor at the centre. We also considered the requirements that each centre had that a tutor must satisfy before tutoring a specific course described in Byerley *et al.* (2020).

4. Results: Data summaries and hypotheses generated with Delphi process

The eight hypotheses from Stage 5 are summarized in Table 1. For example, the cell in Table 1 with a ‘2’ indicates that the hypothesis *Centres with more specialized tutoring models have more visits per student* was the second most popular hypothesis earning 33 points on a scale from −48 to 48. These eight hypotheses were written using ideas from the top three hypotheses from Stage 2. So, it is not surprising that all reported eight were viewed positively, and all scored above zero points when votes were added together. The summary table shows that the leaders tended to be more confident about hypotheses related to how to increase visits per student and return visits than they were about hypotheses related to larger effects of visits on grades. There was not a clear trend about which potential characteristic (a more specialized model, office hours at the centre, more tutor training) was hypothesized to have a bigger impact on effectiveness.

The following sections provide the statistical data and diagrams that were used while voting on the hypotheses with a focus on the top 4 hypotheses from Table 1. The top 8 hypotheses fit into the three themes specialist–generalist tutoring spectrum, tutor training and holding office hours at centre. The charts related to Hypotheses 5 to 8 are in Supplemental Fig. 1, Supplemental Fig. 2, Supplemental Fig. 3 and Supplemental Fig. 4.

4.1 *Quantitative measures of effectiveness*

Table 2 displays the effects of visits to the centre on course grades after controlling for students' university entrance exam scores (USA universities use SAT and ACT equivalent scores) and high school GPA. The R^2 of the model represents the proportion of variance in course grades that can be accounted for by variance in the explanatory variables. For five centres, the model suggests that a higher number of centre visits is a statistically significant predictor of higher grades. The model for Whale, for example, predicts that if students visit the centre 10 times in a semester, their course grade point would be 0.16 higher than students with similar high school GPA and exams scores who did not attend the centre. An increase in course grade point of 1 is the same as an increase in one letter grade. Most commonly in the USA, 4.0 is the highest possible grade point, 2.0 is considered average and 0.0 represents a failure to earn credit.

The data in Table 2 are intended to illustrate data typical tutoring centres may have available, indicate what analyses centre leaders may want to conduct for internal information or external reporting and serve as one source of data for the generation of hypotheses reported in this paper. Pragmatically, it has been extremely useful to share Table 2 with centre leaders and administrators who read annual evaluation reports because it helps them calibrate expectations for how big the predicted increase in grade per one visit might be. At the centre with the largest predicted increase in grade point per one visit (0.035), a student would need to visit roughly twice a week to have a predicted increase in semester grade of 10%.

There are limitations to the conclusions that can be drawn from Table 2. The statistical results presented here are not intended to demonstrate that there is a positive or negative causal relationship at any centre between visits and grades. Our model only partially accounts for self-selection bias and could be improved by adding more predictors, such as student scores on the first exam of the course. In particular, we do not think that attending the centre at Goat and Dolphin *caused* students to earn lower grades, even though there was a statistically significant negative relationship between visits and grades. Instead, we hypothesize that the students who were struggling the most at those schools were more likely to seek tutoring. Dolphin's centre leader reported that instructors primarily encouraged students with low early exam scores to attend the centre, while at other universities all students were encouraged to go. Future studies could use first exam scores as a control variable to investigate negative correlations between visits and grades at centres. We did not do that analysis in this paper because only some of the centre leaders had access to examination grades. However, Horse did have access to examination grades and found a positive correlation between visits and grades when controlling for first exam scores. Finally, the R^2 values were somewhat lower than those in other studies of predictors of college grades but not alarmingly lower (Cohn *et al.*, 2004).

4.2 *Specialist–generalist tutoring spectrum*

This section provides evidence for the hypotheses ranked first and second in Stage 5. These two were both related to the specialist–generalist tutoring spectrum. This section discusses the three sources of evidence used to vote: the diagrams in Figs. 2 and 3 that we created to organize data around the hypotheses, research literature and professional experiences.

Of the eight hypotheses in Stage 5, the top ranked hypothesis (34 points) was: *the more courses a tutor is responsible for tutoring the more likely the tutor will struggle to answer student questions, when the difficulty of the courses being tutored is roughly the same for the tutor tutoring many courses and the tutor tutoring only one course.* Centre leaders self-reported information about the number of courses each tutor tutors and how often the tutor struggles to answer questions (Fig. 2). Leaders typically know

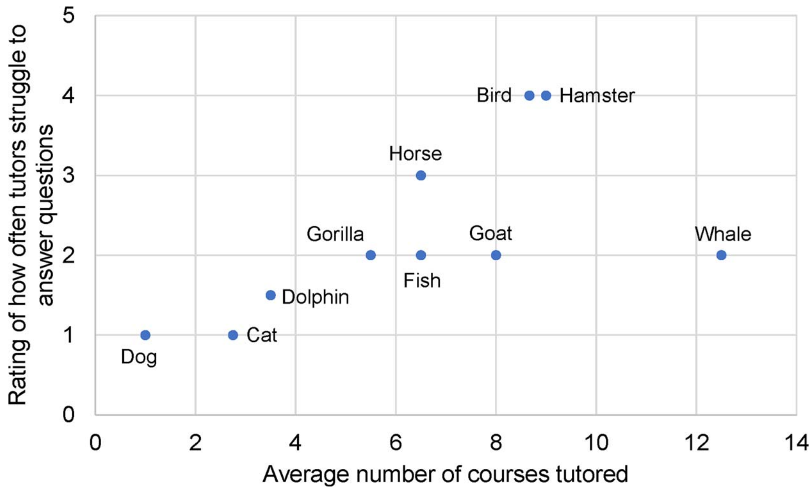

FIG. 2. How often tutors struggle to answer questions compared to the average number of courses tutored by each tutor.

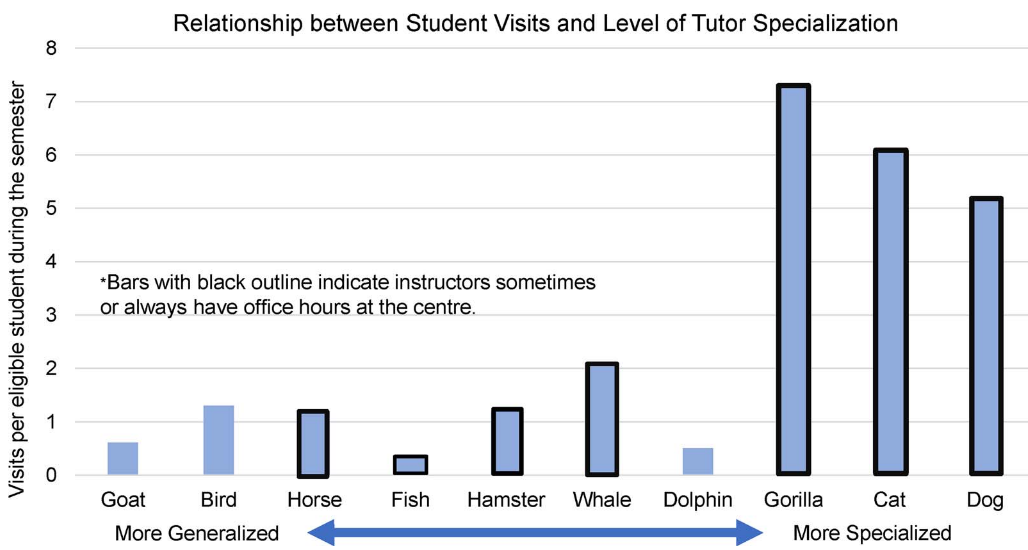

FIG. 3. The mean number of visits per eligible student to the tutoring centre per semester. Eligible students include all students taking a class the centre serves.

how often tutors struggle with questions because the students complain or seek additional help from the leader. The y-axis rates how often tutors struggle to answer questions with 4, every day centre is open; 3, often; 2, occasionally; 1, rarely. The positive relationship displayed in the figure provides support of this hypothesis. The hypothesis includes the nuanced statement 'when the difficulty of the courses being tutored is roughly the same' because in our professional experience it is much easier to tutor multiple lower-level courses than it is to tutor multiple higher-level courses. We had some additional data in the qualitative descriptions about the most difficult courses the centre tutored, which we considered when voting that does not appear on the diagram. Centre leaders were not necessarily concerned when tutors struggled to answer student questions because tutors were often able to eventually solve the problem using resources, and some felt it was beneficial that tutors modelled a problem-solving process. Some students complain when tutors struggle to solve questions, and others appreciate sharing the struggle with someone else. For instance, the mean number of return visits for a student who visited at least once at Bird was very high (14 visits in one semester) despite tutors' frequent difficulties in solving problems.

The second-ranked hypothesis (33 points) was: *Centres with more specialized tutor models have more visits per student.*

Figure 3 shows how we organized both qualitative and quantitative data to investigate this hypothesis. Figure 3 shows that the centres with the three most specialized models also had the most visits per eligible student in a semester. Specialization might increase the number of visits because students are more satisfied with tutors who are more familiar with the course they are taking. However, specialized tutors are often directly involved in students' courses as instructors, learning assistants or graders and that involvement helped them gain specialized knowledge. While working with students in courses, specialized tutors often encourage students to visit the centre. For example, undergraduate tutors at Dog attended courses to help with group work and individually invited students to the centre. It is difficult to disentangle the impact of a specialized structure and the impact of holding office hours in the centre; so, we consider a hypothesis related to office hours in Section 4.4 and include black outlined bars indicated which universities used the practice of having instructors' office hours in the centre in Fig. 3. The centres with the most visits per eligible student all held office hours in the centre, but this did not appear to be a sufficient condition to increase visitation as Horse, Fish, Hamster and Whale had office hours in the centre but substantially fewer visits than Gorilla, Cat and Dog.

In addition to the quantitative data organized in the diagrams, our team used both personal experiences and research to justify the hypotheses about the specialist–generalist tutoring spectrum. All experts provided citations to justify their votes and the first and second authors read and synthesized the suggested research in the following section to provide evidence of the benefits of specialization. Experts justified hypotheses about the benefits of specialization using research showing that it is intellectually challenging to develop the knowledge needed to respond to and support students' mathematical thinking.

Research on mathematics teaching shows that taking mathematics courses (or even earning a mathematics degree) is not sufficient preparation for conveying productive mathematical meanings to students (Byerley *et al.*, 2016; Byerley & Thompson, 2017; Thompson *et al.*, 2017). Further, instructors struggle to attend to and make use of student thinking (Wallach & Even, 2005; Speer & Wagner, 2009; Johnson & Larsen, 2012). It is difficult to respond to a student's suggestions and ideas while tutoring (Arcavi & Schoenfeld, 1992). Speer & Wagner (2009) studied a mathematics professor with 17 years of experience who struggled with adapting instruction in the moment to take into account student contributions. The first time an instructor encounters a specific student difficulty, the teacher must understand the student, unpack the mathematical ideas connected to the misunderstanding and figure out how to connect the student's ways of knowing to conventional mathematics. The next time a teacher encounters a similar misconception, they can more quickly 'call up the ideas she has generated in similar situations as a "chunk" of knowledge' (Speer & Wagner, 2009, p. 559). Specialist undergraduate tutors can build 'chunks' of knowledge (Johns, 2020). The specialist tutor she studied 'engaged in strategies such as scaffolding, converging on shared meaning, error diagnosing and addressing motivation' (Johns, 2020).

Additionally, research on tutoring shows that effective tutors draw upon more than content knowledge and pedagogical knowledge and have developed additional insight into learning mathematics

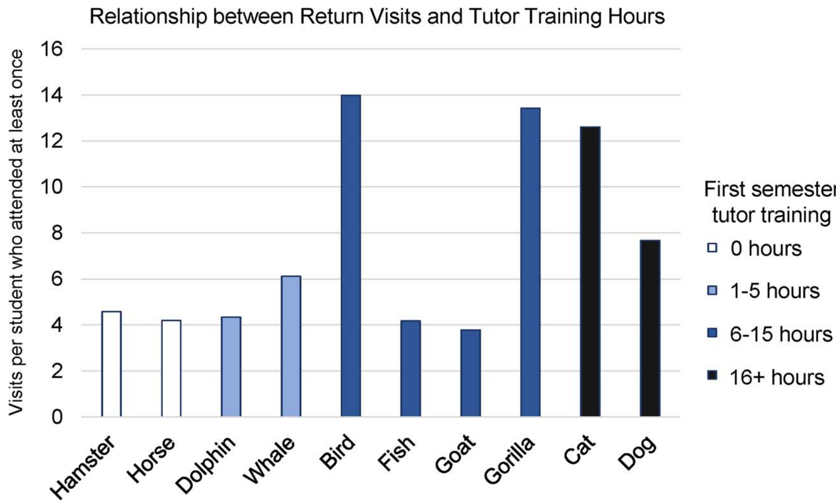

FIG. 4. The centres are organized from those requiring the least to most training.

(Lepper & Woolverton, 2002). For example, expert tutors in Lepper and Woolverton's study had both subject-specific content knowledge and the ability to form a 'cognitive model that is focused on the student's current state of knowledge' (p. 142). They describe the most effective tutors as providing historical information motivating to students, visual models and real-world analogies. The best tutors knew 'what sorts of problems were most likely to prove especially difficult for students or to elicit particular sorts of errors from them' (Lepper & Woolverton, 2002). Tutors also use knowledge of content and curriculum (Ball *et al.*, 2008) which entails knowledge about course sequencing, course goals, materials available for instruction and course relationship to other courses. Knowledge used while tutoring also includes knowledge about exams, online homework systems and course websites (Johns & Burks, 2022).

The voting scores indicate we are more confident that a specialized structure increases the number of visits per student than student grades. While generalist tutors cannot always immediately solve a student's problem, some of us saw value in tutors modelling the problem-solving process including productive struggle and effective use of resources such as textbooks and notes. Future research investigating specialized versus generalized structures could include videotaping tutors working in courses of specialization and courses with which they are less familiar to investigate the effectiveness of each session from the perspective of the student. Centre leaders could also modify the organizational structure of their centre and compare measures of effectiveness from one semester to the next.

4.3 *Tutor training hypotheses*

We hypothesized that tutor training was positively linked to more return visits per student and a larger impact of tutoring on grades. We voted for the third ranked hypotheses in Stage 5 using professional experience, research and the data organized in Fig. 4. We counted the tutors' concurrent experiences as learning assistants or instructors as training hours.

The third-ranked hypothesis (25 points) was: *Centres with more training for tutors have more return visits per student when tutor training is high-quality and meets the needs of tutors.*

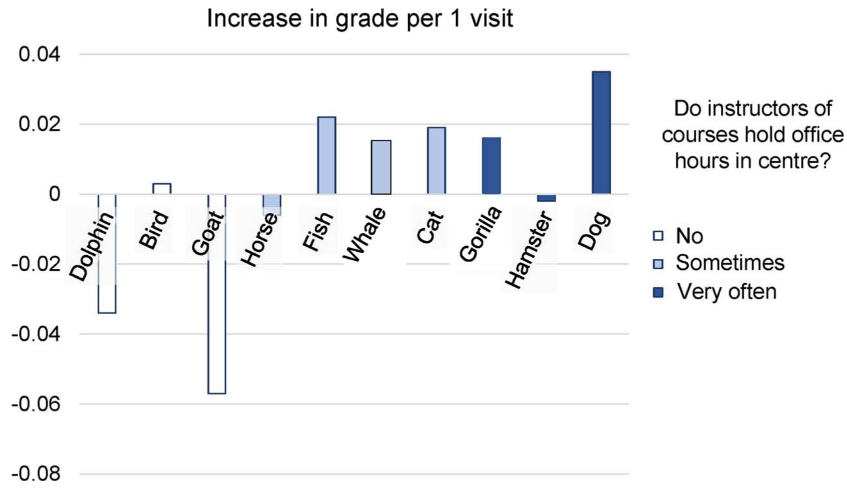

FIG. 5. Centres that held instructor office hours at the centre tended to have a larger effect of visits on grades.

The initial hypotheses generated about training in Stage 1 suggested that tutor training strengthens a centre. After reflection on professional experience, we noticed that if the training is not high quality and suited to the needs of the tutor it is unlikely to improve their effectiveness. As justification one leader noted that teacher training does not always positively impact student success (Harris & Sass, 2011). We did not have the means to evaluate the quality of the training at each of the 10 centres in comparison to each other but felt the quality of training mattered. We added the caveat that the training needs to meet the needs of the tutors based on observations that some tutors need more training in content and others need more training in pedagogy. For example, one centre leader wrote about graduate students who provided high-level mathematical justifications to undergraduate students using ideas and terminology from their graduate courses. Those graduate tutors had a deep understanding of the mathematics but needed training in using language and ideas accessible to the students they were tutoring. In contrast, Goat's tutors often tutor courses they have never taken such as 'Logic, Set Theory and Probability.' The STEM majors at Goat who were most interested in being tutors were not required to take courses covering the mathematical content in the required courses for non-STEM majors. Thus, although Goat's centre had a relative high number of training hours it might be that the subject-specific training they received from the centre could not fully remedy the institutional structure that resulted them tutoring mathematical content they had not been asked to learn in a course.

Based on professional experience and literature on teacher training (van Es & Sherin, 2010), we believe training can help tutors develop student-centred pedagogical tools, such as questioning and listening practices. Training can also help tutors be aware of how small behaviours can send positive or negative signals to students (MacGillivray & Croft, 2009). Johns', 2020 observational study of tutors, as well as our collective experience, suggests that tutors are able to change their practices to promote more student engagement. We have had success asking tutors to follow Lepper and Woolverton (2002) suggestion to give students up to five or six hints and wait patiently before directly telling students how to solve a problem (Lepper & Woolverton, 2002). Other studies about tutoring commonly promote active inquiry and self-explanations on the part of the student (Chi, 1996; Lepper & Woolverton, 2002; Topping, 2005), as well as appropriate questioning and responsive scaffolding on the part of the tutor

(Topping, 1996; Roscoe & Chi, 2007; Graesser *et al.*, 2011). We hypothesize that training can help tutors develop these questioning, scaffolding and wait time skills more rapidly than if they are left to learn on their own. Although based on our experience, we believe tutors can be trained to enact effective practices, there is little to no literature about the effectiveness of various types of tutor training to respond to the needs of each centre.

In this study, we did not attempt to evaluate the quality of training offered, nor how much time was spent on pedagogical versus content training as our main purpose was to generate hypotheses. In Byerley *et al.* (2020), hours spent on general tutor training and course-specific training were reported separately, but we grouped all training together in these charts because the generated hypotheses did not distinguish training types. In future research, it would be important to gather data related to what knowledge tutors need and what knowledge is targeted in training.

4.4 Holding office hours in the centre hypotheses

In Stage 5, we hypothesized holding office hours in a centre was positively associated with having larger effects of visits on grades (fourth ranked) and more visits per eligible student (fifth ranked) (Fig. 5). The experts found office hours to be the most important one related to the strength of the relationship between the centre and the mathematics department.

The fourth-ranked hypothesis (16 points) was: *Centres where instructors hold office hours in the centre have larger positive effects of visits on grades when instructors are not resentful or annoyed about being asked to work in the centre*.

Our sources of evidence for these hypotheses are primarily the quantitative data and our professional experience. In our quantitative data, we saw Cat, Dog and Gorilla had many office hours at their centres, a high number of visits per student and above average scores on other metrics of success. Receiving tutoring from an instructor, or a tutor who can easily talk to an instructor at the centre, might be more effective because the tutors are more likely to help students develop knowledge specifically needed for the course. Based on professional experience, we hypothesize that there are a number of benefits of instructors holding office hours in the centre with the caveat that this is effective 'when instructors are not resentful or annoyed about being asked to work in the centre.' For instance, in the qualitative descriptions the experts read, Hamster's leader reported that the tutors and instructors at the centre often ignore students and play on their phones. Hamster's leader personally observed this and suggested the importance of adding the clause about the instructor's attitudes about tutoring to the hypotheses during the Delphi process. More positively, students likely feel more comfortable going to a centre where they can sit, study and ask questions when needed than they would at an instructor's personal office. Instructors who hold office hours at the centre are also likely to advertise to their students the location of the centre which should lead to increase visitation and a potential to build a community at the centre.

4.5 Data related to unpopular hypotheses

There were centre attributes for which we collected data that we were unable to generate widely agreed on hypotheses, such as tutoring capacity, centre size, centre location, job title of centre leader and type of service. There was wide variation in the amount and quality of space available per student visit but we did not see patterns between the quality of space and measures of effectiveness. Area per eligible student ranged from 0.019 to 0.19 m^2 (see Byerley *et al.* (2020) for the area of each centre.) We did not see patterns relating the amount of time tutors could spend per student and measures of effectiveness such as correlations between visits and grades. Tutor time available per student visit ranged from a mean of 11 to 80 min.

5. Conclusions and limitations

The biggest strength of this work is that we provide concrete hypotheses on how to improve the effectiveness of a centre that are feasible to implement in practice. These suggestions are based on the experience of multiple experienced centre leaders in multiple contexts and are supported by exploratory data analysis. The primary limitation of this research is that we collected the data and designed the methods to generate hypotheses, not to test them. Further, although we had a sample size of over 26,000 students at 10 institutions, it was a convenience sample of 10 centres from only the USA. While we would certainly feel comfortable recommending that other centre leaders increase training time, hold instructor office hours in their centre or ask tutors to specialize in a small number of classes, we would not argue that our methods completely validate our hypotheses. To test the hypotheses would require research designed specifically for this purpose. For example, a tutoring centre might reduce the number of courses each tutor is responsible for tutoring for a semester while keeping other policies constant. For example, as a result of participating in this study Bird implemented specialized tutoring in several courses (statistics and business mathematics) and saw increases in student visits for those two courses.

Further, we formed our hypotheses on the assumption that all effective centres would have the same structure and used the same measures of effectiveness across all centres. However, organization literature asserts that there is no one best organizational structure and instead should be based on local context (Donaldson, 1996). Similarly, organizational literature suggests that effectiveness measures should also be based on the local context, and a good measure of effectiveness at one centre should not necessarily be used to measure another. As a follow-up to this study, Johns *et al.* (2021) explored applying organizational literature to examining the effectiveness of centres. Later, in our self-study, we learned of Weisbord's (1976) Six-Box model for describing organizational structure and identified other dimensions of organizational structure that are potentially related to the effectiveness of centres. Examples include rewards, such as course buy-outs or pay for centre leaders, mechanisms, such as queuing systems, and leadership, such as how centre leaders monitor and improve the centre. Finally, the COVID-19 pandemic occurred after the data collection, but future research should consider organizational structures involving remote access as described in Gilbert *et al.* (2021).

Despite the acknowledgement of our limitations, we argue this research program has enormous potential to improve students' mathematics tutoring experiences. Centres are a key resource to support students' success in mathematics, and it is important to use research and collaboration to make them as effective as possible. Based on our experience, tutor centre leaders typically have the power to make substantial changes to the structure of their centres, and some changes could be implemented that require little in the way of financial resources. Through the development of hypotheses on successful tutoring centre structures, we have a strong foundation for further investigation into characteristics of effective centres and have provided current practitioners ideas to consider incorporating and templates of ways they might organize their data.

Acknowledgements

We are grateful for the National Science Foundation conference grant (DUE: 2645086) for tutoring centre leaders. Collaboration at those conferences was the inspiration for this paper. The opinions expressed in this paper do not necessarily reflect the position, policy, or endorsement of the supporting agency.

References

Arcavi, A. & Schoenfeld, A. H. (1992) Mathematics tutoring through a constructivist lens: the challenges of sense-making. *J. Math. Behav.*, 11, 321–335.

Ball, D. L., Thames, M. H. & Phelps, G. (2008) Content knowledge for teaching what makes it special? *J. Teach. Educ.*, 59, 389–407.

Berkopes, K. & Abshire, S. (2016) Quantitative measures for assessing learning Centers: An agenda and exploration. *Learning Assistance Review*, 21, 109–126.

Berry, E., Mac An Bhaird, C. & O'Shea, A. (2015) Investigating relationships between the usage of mathematics learning support and performance of at-risk students. *Teaching Mathematics Applications*, 34, 194–204.

Brady, S. R. (2015) Utilizing and adapting the Delphi method for use in qualitative research. *Int J Qual Methods*, 14, 1–6, 160940691562138.

Burks, L. & James, C. (2019) Mathematical Knowledge for Tutoring Undergraduate Mathematics. *Proceedings of the 22nd Annual Conference on Research in Undergraduate Mathematics Education* (A. Weinberg, D. Moore-Russo, H. Soto & M. Wawro eds). OK: Oklahoma City, pp. 731–740.

Byerley, C. & Thompson, P. W. (2017) Secondary mathematics teachers' meanings for measure, slope, and rate of change. *J. Math. Behav.*, 48, 168–193.

Byerley, C., Yoon, H. & Thompson, P. W. (2016) Limitations of a "chunky" meaning for slope. *Proceedings of the 19th Annual Conference on Research in Undergraduate Mathematics Education*. PA: Pittsburgh.

Byerley, C., Campbell, T. & Rickard, B. (2018) Evaluation of Impact of Calculus Center on Student Achievement. *Proceedings of the 21st Annual Conference on Research in Undergraduate Mathematics Education* (A. Weinberg, C. Rasmussen, J. M. Rabin, M. Wawro & S. Brown eds). CA: San Diego, pp. 816–825.

Byerley, C., Moore-Russo, D., James, C., Johns, C., Rickard, B. & Mills, M. (2019) Defining the varied structures of tutoring centers: Laying the foundation for future research. *Proceedings of the 22nd Annual Conference on Research in Undergraduate Mathematics Education* (A. Weinberg, D. Moore-Russo, H. Soto & M. Wawro eds). OK: Oklahoma City, pp. 93–101.

Byerley, C., James, C., Moore-Russo, D., Rickard, B., Mills, M., Heasom, W., Oien, J., Farthing, C., Burks, L., Ferreira, M., Mammo, B. & Moritz, D. (2020) Characteristics and Evaluation of Ten Tutoring Centers. *Proceedings of the 23rd Annual Conference on Research in Undergraduate Mathematics Education* (S. Karunakaran, Z. Reed & A. Higgins eds). MA: Boston, pp. 70–78.

Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V. & Hiebert, J. (2019) Research pathways that connect research and practice. *J. Res. Math. Educ.*, 50, 2–10. https://doi.org/10.5951/jresematheduc.50.1.0002.

Chi, M. T. (1996) Constructing self-explanations and scaffolded explanations in tutoring. *Appl. Cogn. Psychol.*, 10, 33–49.

Clayton, M. J. (1997) Delphi: a technique to harness expert opinion for critical decision-making tasks in education. *Educ. Psychol.*, 17, 373–386.

Cohen, J., Cohen, P., West, S. G. & Aiken, L. S. (2013) *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Routledge.

Cohn, E., Cohn, S., Balch, D. C. & Bradley Jr., J. (2004) Determinants of undergraduate GPAs: SAT scores, high-school GPA and high-school rank. *Econ. Educ. Rev.*, 23, 577–586.

Croft, T. & Grove, M. (2016) Mathematics and statistics support centres: resources for training postgraduates and others who work in them. *MSOR Connections*, 14, 3–13.

Cronin, A. & Meehan, M. (2021) The utility of mathematics support Centre feedback to lecturers of large first-year university mathematics courses. *Int. J. Math. Educ. Sci. Technol.*, 52, 1472–1490.

Cronin, A., Cole, J., Clancy, M., Breen, C. & O'Sé, D. (2016) An audit of Mathematics Learning Support provision on the island of Ireland in 2015. *National Forum for the Enhancement of Teaching and Learning in Higher Education*. The National Forum for the Enhancement of Teaching and Learning in Higher Education.

Delderfield, R. & McHattie, H. (2018) The person-centred approach in maths skills development: examining a case of good practice. *Journal of Learning Development in High. Educ.*, 13.

Donaldson, L. (1996) *For positivist organization theory: Proving the hard core*. London: Sage Publications Ltd.
Ellis, J., Fosdick, B. K. & Rasmussen, C. (2016) Women 1.5 times more likely to leave STEM pipeline after calculus compared to men: lack of mathematical confidence a potential culprit. *PLoS One*, 11, e0157447.
van Es, E. A. & Sherin, M. G. (2010) The influence of video clubs on teachers' thinking and practice. *J. Math. Teach. Educ.*, 13, 155–176.
Fitzmaurice, O., Cronin, A. G., Ni Fhloinn, E., O'Sullivan, C. & Walsh, R. (2016) Preparing tutors for mathematics learning support. *MSOR Connections*, 14, 14–21.
Fletcher, A. J. & Marchildon, G. P. (2014) Using the Delphi method for qualitative, participatory action research in health leadership. *Int J Qual Methods*, 13, 1–18.
Gilbert, H., Hodds, M. & Lawson, D. (2021) 'Everyone seems to be agreeing at the minute that face-to-face is the way forward': practitioners' perspectives on post-pandemic mathematics and statistics support. *Teaching Mathematics and its Applications: An International Journal of the IMA*, 40, 296–316.
Gillard, J., Robathan, K. & Wilson, R. (2011) Measuring the effectiveness of a mathematics support service: an email survey. *Teaching Mathematics and Its Applications: International Journal of the IMA*, 30, 43–52.
Gioia, D., Patvardhan, S., Hamilton, A. & Corley, K. (2013) Organizational identity formation and change. *Acad. Manag. Ann.*, 7, 123–193. https://doi.org/10.5465/19416520.2013.762225.
Goertzen, R. M., Brewe, E., Kramer, L. H., Wells, L. & Jones, D. (2011) Moving toward change: institutionalizing reform through implementation of the learning assistant model and open source tutorials. *Physical Review Special Topics-Physics Education Research*, 7.
Gordon, N. (2004) Mathematics and computing. *MSOR Connections*, 4, 10–13.
Graesser, A. C. & Person, N. K. (1994) Question asking during tutoring. *Am. Educ. Res. J.*, 31, 104–137.
Graesser, A. C., Rus, V. & Hu, X. (2011) Instruction based on tutoring. *Handbook of Research on Learning and Instruction* (R. E. Mayer & P. A. Alexander eds). New York: Routledge, pp. 408–426.
Grove, M., Croft, T., Lawson, D. & Petrie, M. (2018) Community perspectives of mathematics and statistics support in higher education: building the infrastructure. *Teaching Mathematics and its Applications: An International Journal of the IMA*, 37, 171–191.
Grove, M., Croft, T., Lawson, D. & Petrie, M. (2019) Community perspectives of mathematics and statistics support in higher education: the role of the staff member. *Teaching Mathematics and its Applications: An International Journal of the IMA*, 38, 43–59.
Grove, M., Croft, T. & Lawson, D. (2020) The extent and uptake of mathematics support in higher education: results from the 2018 survey. *Teaching Mathematics and its Applications: An International Journal of the IMA*, 39, 86–104.
Gutiérrez, R. (2013) The sociopolitical turn in mathematics education. *J. Res. Math. Educ.*, 44, 37–68.
Harris, D. N. & Sass, T. R. (2011) Teacher training, teacher quality and student achievement. *J. Public Econ.*, 95, 798–812.
Ireland, L. (2006) Maths support at the University of Hull: what we do and what we have learned. *MSOR Connections*, 6, 1–3.
Jacob, M. & Ní Fhloinn, E. (2019) A quantitative, longitudinal analysis of the impact of mathematics support in an Irish university. *Teaching Mathematics and its Applications: An International Journal of the IMA*, 38, 216–229.
Johns, C. (2020) Tutoring beyond show and tell: An existence proof. *Proceedings of the 23rd Annual Conference on Research in Mathematics Education* (S. Karunakaran, Z. Reed & A. Higgins eds). MA: Boston, pp. 294–301.
Johns, C. A. & Burks, L. C. (2022) A framework for mathematical knowledge for undergraduate mathematics tutors. *International Journal of Research in Undergraduate Mathematics Education*, 1–30.
Johns, C., Byerley, C., Moore-Russo, D., Rickard, B., Oien, J., Burks, L., James, C., Mills, M., Heasom, W., Ferreira, M. & Mammo, B. (2021) Performance assessment for mathematics tutoring centres. *Teaching Mathematics and its Applications: An International Journal of the IMA*.
Johnson, E. & Hanson, K. (2015) *Academic and social supports. In: Insights and recommendations from the MAA*

national study of college calculus (D. Bressoud, V. Mesa & C. Rasmussen eds). Washington, DC: MAA Press, pp. 69–92.

Johnson, E. & Larsen, S. P. (2012) Teacher listening: the role of knowledge of content and students. *J. Math. Behav.*, 31, 117–129.

Lawson, D., Grove, M. & Croft, T. (2020) The evolution of mathematics support: a literature review. *Int. J. Math. Educ. Sci. Technol.*, 51, 1224–1254.

Lepper, M. R. & Woolverton, M. (2002) The wisdom of practice: Lessons learned from the study of highly effective tutors. *Improving academic achievement: Impact of psychological factors on education* (J. Aronson ed). Amsterdam: Academic Press, pp. 135–158.

Linstone, H. A. & Turoff, M. (1975) *The Delphi method: Techniques and applications*. Reading, MA: Addison-Wesley Pub. Co.

MacBeath, J., Kirwan, T., Myers, K., McCall, J., Smith, I., McKay, E., Sharp, C., Bhabra, S., Weindling, D. & Pocklington, K. (2001) *The impact of study support: A report of a longitudinal study into the impact of participation in out-of-school-hours learning on the academic attainment, attitudes and school attendance of secondary school students*. Colegate, Norwich: Department for Education and Skills.

MacGillivray, H. & Croft, T. (2009) Learning support and students studying mathematics and statistics. *Int. J. Math. Educ. Sci. Technol.*, 40, 455–472.

MacGillivray, H. & Croft, T. (2011) Understanding evaluation of learning support in mathematics and statistics. *Int. J. Math. Educ. Sci. Technol.*, 42, 189–212.

Manizade, A. G. & Mason, M. M. (2011) Using Delphi methodology to design assessments of teachers' pedagogical content knowledge. *Educ. Stud. Math.*, 76, 183–207.

Matthews, J., Croft, T., Lawson, D. & Waller, D. (2013) Evaluation of mathematics support centres: a literature review. *Teaching Mathematics and its Applications: International Journal of the IMA*, 32, 173–190.

McKenna, H. P. (1994) The Delphi technique: a worthwhile research approach for nursing? *J. Adv. Nurs.*, 19, 1221–1225. https://doi.org/10.1111/j.1365-2648.1994.tb01207.x.

Mills, M., Tallman, M. & Rickard, B. (2017) Research opportunities for RUME researchers in the context of mathematics resource centers. *Working group at the 20th Annual Conference on Research in Undergraduate Mathematics Education*. San Diego, CA.

Mills, M., Rickard, B. & Guest, B. (2020) Survey of mathematics tutoring centres in the USA. *Int. J. Math. Educ. Sci. Technol.*, 53, 1–21.

Moore-Russo, D., Tinsley, C. & Brady, N. (2018) Studying a math help center using organizational development and change theory. *Presentation at the Transforming STEM Higher Education Conference*. Atlanta, GA: Association of American Colleges and Universities with Project Kaleidoscope.

Mullen, C., Cronin, A., Taylor, L. & Lui, C. (2021) Evaluating the impact of mathematics support using moderation. *Eighth Conference on Research in Mathematics Education in Ireland* (M. Kingston & P. Grimes eds). Dublin City University, pp. 284–291.

Muñiz-Rodríguez, L., Alonso, P., Rodríguez-Muñiz, L. J. & Valcke, M. (2017) Developing and validating a competence framework for secondary mathematics student teachers through a Delphi method. *J. Educ. Teach.*, 43, 1–17.

Pill, J. (1971) The Delphi method: substance, context, a critique and an annotated bibliography. *Socio Econ. Plan. Sci.*, 5, 57–71.

Rickard, B. & Mills, M. (2018) The effect of attending tutoring on course grades in calculus I. *Int. J. Math. Educ. Sci. Technol.*, 49, 341–354.

Roscoe, R. D. & Chi, M. T. (2007) Understanding tutor learning: knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Rev. Educ. Res.*, 77, 534–574.

Ryals, M., Johns, C. & Mills, M. (2019) Undergraduate mathematics tutors and students' challenges of knowing-to act. *Proceedings of the 22nd Annual Conference on Research in Undergraduate Mathematics Education* (A. Weinberg, D. Moore-Russo, H. Soto & M. Wawro eds). OK: Oklahoma City, pp. 1017–1022.

Rylands, L. J. & Shearman, D. (2018) Mathematics learning support and engagement in first year engineering. *Int. J. Math. Educ. Sci. Technol.*, 49, 1133–1147.

Schoenfeld, A. H., Gamoran, M., Kessel, C. & Leonard, M. (1992) Toward a comprehensive model of human tutoring in complex subject matter domains. *J. Math. Behav.*, 11, 293–319.

Skulmoski, G. J., Hartman, F. T. & Krahn, J. (2007) The Delphi method for graduate research. *Journal of Information Technology Education: Research*, 6, 001–021.

Speer, N. M. & Wagner, J. F. (2009) Knowledge needed by a teacher to provide analytic scaffolding during undergraduate mathematics classroom discussions. *J. Res. Math. Educ.*, 40, 530–562.

Starkings, S. (2002) Mathematical and statistical support service at south Bank University. *MSOR Connections*, 2, 13–15.

Tallman, M. A., Carlson, M. P., Bressoud, D. M. & Pearson, M. (2016) A characterization of calculus I final exams in US colleges and universities. *International Journal of Research in Undergraduate Mathematics Education*, 2, 105–133.

Thompson, P. W., Hatfield, N. J., Yoon, H., Joshua, S. & Byerley, C. (2017) Covariational reasoning among US and south Korean secondary mathematics teachers. *J. Math. Behav.*, 48, 95–111.

Topping, K. J. (1996) The effectiveness of peer tutoring in further and higher education: a typology and review of the literature. *High. Educ.*, 32, 321–345.

Topping, K. J. (2005) Trends in peer learning. *Educ. Psychol.*, 25, 631–645.

Walker, K. & Dancy, M. (2006) Investigation and Evaluation of a Physics Tutorial Center. *Paper presented at Physics Education Research Conference 2006*. Syracuse, New York: Retrieved December 12, 2022, from https://www.compadre.org/Repository/document/ServeFile.cfm?ID=5268&DocID=3505.

Wallach, T. & Even, R. (2005) Hearing students: the complexity of understanding what they are saying, showing, and doing. *J. Math. Teach. Educ.*, 8, 393–417.

Walsh, R. (2017) A case study of pedagogy of mathematics support tutors without a background in mathematics education. *Int. J. Math. Educ. Sci. Technol.*, 48, 67–82.

Weisbord, M. R. (1976) Organizational diagnosis: six places to look for trouble with or without a theory. *Group & Organization Studies*, 1, 430–447.

Xu, Y., Hartman, S., Uribe, G. & Mencke, R. (2001) The effects of peer tutoring on undergraduate students' final examination scores in mathematics. *Journal of College Reading and Learning*, 32, 22–31.

Cameron Byerley is an assistant professor at the University of Georgia. Formally, she co-directed a university mathematics tutoring center. Her research interests include modeling students' and citizens' mathematical thinking.

Carolyn Johns is the assistant director of the Mathematics and Statistics Learning Center at Ohio State. Her research interests broadly include mathematics tutor practices, tutor training and teacher education. Prior to working at The Ohio State University, she received an undergraduate degree from Earlham College and received a Ph.D. from The Ohio State University in August 2019.

Deborah Moore-Russo, Ph.D. is the First-Year Mathematics Director at the University of Oklahoma. Her primary research interests include the ways that mathematics topics are conceptualized, represented, visualized and communicated. She is also interested in how mathematics departments support first-year mathematics courses through tutoring centers, course coordination and other efforts.

Brian Rickard earned a BS in Mathematics, MEd in Higher Education and PhD in Educational Statistics and Research Methods at the University of Arkansas. He is a teaching assistant professor and coordinator of finite mathematics in the Department of Mathematical Sciences at the University of Arkansas. He conducts research in undergraduate mathematics education and university tutoring centers.

Carolyn James, PhD, is the Calculus Coordinator and Mathematics Resource Center supervisor at University of Portland. She received her undergraduate degree in mathematics at Carleton College and her PhD in mathematics education from Portland State University. Her research interests include barriers and drivers of instructional change, the relationship between instructor beliefs and practice, and undergraduate mathematics tutoring.

Melissa Mills is a teaching associate professor and the Director of the Mathematics Learning Success Center at Oklahoma State University. Her research interests broadly include mathematics tutor practices and the teaching and learning of mathematical proof at the undergraduate level.

Behailu Mammo is a professor of mathematics and Director of the Mathematics Tutorial Center at Hofstra University, NY. His research interest focuses on number theory and examining the impact of tutoring in improving pre-service teachers' dispositions and self-efficacy. In May 2005, he received his Ph.D. from Temple University, PA. His bachelor's and master's degrees are from Ethiopia, East Africa.

Janet Oien is the Co-Director of the Calculus Center at Colorado State University. Her research interests broadly include mathematics tutor practices, tutor training and preparation of secondary mathematics teachers. She received her undergraduate degree and M.S. degree in Mathematics with an outside concentration in Education from Colorado State University in 2006.

Linda Burks is the Director of the Mathematics Learning Center and a Math Lecturer at Santa Clara University. She received her undergraduate degree in mathematics from Virginia Tech; she earned her M.S. in mathematics and Ph.D. in mathematics education from the University of Michigan. Her research interests include mathematics tutor practices and cognitive self-regulation.

William Heasom, MA, MSCE, is Assistant to the Director of the Mathematics Learning and Resource Center at Villanova University He also serves as collaborator with the IARPA FOCUS (Forecasting Counterfactuals in Uncontrolled Settings) project at University of Pennsylvania where he helped develop training materials for intelligence analysts. His research interests concern applying mathematical and agent-based modeling techniques to the analysis and design of hydrologic, ecological and learning systems.

Melissa Ferreira is the Director of the Mathematics Learning and Resource Center at Villanova University. Her research interest is in university mathematics tutoring. She received undergraduate and masters degrees in Mathematics from Villanova University.

Cynthia Farthing is an Associate Professor of Instruction in mathematics and former director of the Math Tutorial Lab at the University of Iowa. Her professional interests include improving introductory undergraduate mathematics courses by revising assessment practices and incorporating metacognitive supports for students.

Daniel Moritz is an Academic Specialist and Lecturer at the University of Colorado, Boulder. He is the founder and director of the Mathematics Academic Resource Center and the Departmental Coordinator of the Math Department's Learning Assistant Program.