

The impact of shadow education on student academic achievement: Why the research is inconclusive and what can be done about it

Mark Bray

Published online: 3 May 2014

© Education Research Institute, Seoul National University, Seoul, Korea 2014

Abstract Recent decades have brought global expansion of private supplementary tutoring, widely known as shadow education. Such tutoring consumes considerable resources and is usually viewed by participating households as an investment that will increase the recipients' academic achievements. However, research on the effectiveness of tutoring has delivered inconclusive and even contradictory findings. Part of the reason lies in definitions and foci of research, since private supplementary tutoring may have many different formats, delivery mechanisms and intensities. As with other types of education, it is important also to investigate the qualities of the instructors and the motivations of the students. This creates a complex research arena, in which practical challenges arise from shortcomings in the instruments for securing data and limitations in methods for analyzing the data. This paper assesses the current state of the literature and makes recommendations for the future research agenda.

Keywords Shadow education · Private tutoring · Academic achievement · Research methods

Introduction

Parts of East Asia have long been known for the scale and intensity of private supplementary tutoring (Lee et al. 2010; Rohlen 1980; Zeng 1999). In recent years, private supplementary tutoring has become a global phenomenon (Bray 2009; Lee et al. 2009; Mori and Baker 2010). The

activity consumes considerable household resources, occupies much of the time of children and youths and has far-reaching social and economic implications. In the Republic of Korea, 86.8 % of elementary students and 72.2 % of middle school students were estimated to have received supplementary tutoring in 2010, with the tutoring sector consuming approximately 3 % of Gross Domestic Product (Lee 2013: 41, 52). This was an extreme case, but proportions were high and growing in many other parts of the world (Bray 2009: 18–19).

Within the literature, private supplementary tutoring is widely called shadow education (e.g., Aslam and Atherton 2012; Bray 1999; Buchmann 2002; Lee et al. 2009). The vocabulary is used because much content of private supplementary tutoring imitates that in the schools. If the authorities change the curriculum in the schools, before long it changes in the shadow. However, like many metaphors, the vocabulary has limitations. In particular, some forms of private tutoring supplement schooling with additional content rather than just mimicked content. Issues of vocabulary and focus need attention from a methodological perspective since they can contribute to confusions.

An obvious question for stakeholders is whether private supplementary tutoring “works” in the sense of raising students' academic achievement. It is easy to find strong affirmative answers. Crotty (2012), for example, writing as a journalist, declared that private tutoring “is a highly effective way to ensure academic excellence.” As one might expect, similar sweeping statements are commonly presented by the industry itself (e.g., Growing Stars 2013; Tutors International 2013). Perhaps more surprising are equally sweeping statements from parts of the academic community. Baily (2012: 382) declared that “the benefit of private tutoring is unarguable”; and Ünal et al. (2010: 5513) asserted that “private tutoring, like any teaching and

M. Bray (✉)
UNESCO Chair Professor in Comparative Education, The
University of Hong Kong, Pokfulam, Hong Kong, China
e-mail: mbray@hku.hk

learning interaction, undoubtedly has positive outcomes for individuals.” Such statements need scrutiny. Much must surely depend on the nature of the tutoring, the volumes of tutoring, the qualities of the tutors, the motivations of the students and many other factors. More careful researchers would agree with Byun (2014: 40) that “empirical evidence has been inconsistent, contradictory and even confusing.”

With such matters in mind, this paper reviews literature on the impact of private supplementary tutoring on students’ academic achievement. Some parts of the literature are derived from large international surveys, others are based on large national surveys, and others are smaller quantitative and qualitative assessments. The paper notes the value of these studies, but also observes problems in the ways that data have been collected and interpreted. These factors are among the reasons why the research has delivered inconsistent findings. The paper remarks on the implications of the current state of research in this domain and looks ahead to a future agenda with accompanying strategies.

Defining terms and setting parameters

In line with much existing literature (e.g., Bray 1999; Bregvadze 2012; Lee et al. 2009; Silova 2009; Silova et al. 2006), the definition of private supplementary tutoring adopted by this paper has three components. First, the adjective *private* indicates that the tutoring is provided in exchange for a fee. The paper is not concerned with tutoring provided free of charge by teachers, family members, community bodies, or others. Second, the *supplementary* nature means that the tutoring is provided in addition to regular schooling. Most commonly the tutoring is received in the homes of the tutors or students, or in tutorial centers run by companies. Third, the focus of this paper is on *academic* subjects. Thus, it is concerned with extra mathematics, languages, sciences, etc. as taught in the regular schools. It is not concerned with sports, music and artistic activities designed for more rounded development. Nor is the paper concerned with religious content or minority languages unless the instruction is designed to support school work in those domains.

The elaboration of definitions above is important because different studies have used different definitions. Thus Stevenson and Baker (1992), for example, included in their definition of shadow education in Japan activities for students who had left school but who attended institutions called *yobiko* which provided support for re-taking examinations. A similar focus was adopted more recently by Coniam (2013) in Hong Kong. Also, some authors (e.g.,

Baker et al. 2001) have included provision of fee-free extra lessons as part of their definition of shadow education; and others (e.g., Xue and Ding 2009) have included non-academic subjects alongside academic ones.

The next question concerns the definition of tutoring. For many people, this term means one-to-one or perhaps small-group instruction (e.g., Medway 1995; Mischo and Haag 2002). In the domain of shadow education, it may include large classes and even full lecture-theaters. In some societies, the term also embraces classes with video replay rather than live tutors (Kwo and Bray 2011); and increasingly tutoring is provided over the internet either live or in recorded form (Ventura and Jang 2010).

A further question concerns the duration and intensity of tutoring. Unlike regular schooling, in which students are assumed to attend lessons 5 days a week during term time, tutoring may be received on highly variable schedules according to demand and supply. Some students receive multiple tutorial lessons regularly each week, while others receive only occasional lessons when they have a particular need. Variations are common at different points in the school year and between different grades.

A further conceptual and measurement challenge concerns student achievement. The present paper is concerned with achievements in academic subjects that are measured through cognitive tests of various kinds. It is not concerned with such matters as self-image, persistence, curiosity and confidence, which may also be important outcomes of tutoring. Nevertheless, even this restricted focus encounters problems of measurement; and when learning gains have been achieved, it may be difficult to ascribe them unambiguously to private tutoring.

Finally, shadow education may subtract as well as supplement. In some countries, teachers commonly provide private tutoring for pupils for whom they are already responsible in regular classes (see e.g., Bray 2009: 80–81; Dawson 2010: 20). In such circumstances, teachers may be tempted to reduce the effort and curriculum coverage of their regular classes in order to promote demand for their private services. Elsewhere teachers are prohibited from providing private supplementary tutoring, but if tutoring is widespread the teachers may assume that their pupils have back-up support and therefore put less effort into their work than they would otherwise.

Large international surveys

Among large international studies of educational achievement, two have become particularly well known. One is the Trends in International Mathematics and Science Study (TIMSS) operated under the auspices of the International Association for the Evaluation of Educational

Achievement (IEA);¹ and the other is the Programme for International Student Assessment (PISA) operated under the auspices of the Organisation for Economic Co-operation and Development. TIMSS has assessed the academic achievements of students in Grades 4 and 8 in mathematics and science every 4 years since 1995, and PISA has assessed the academic achievements of 15-year-olds, most of whom are in Grade 9, every 3 years since 2000.

TIMSS has developed powerful datasets for analysis. However, for research on the impact of private supplementary tutoring, TIMSS has had shortcomings, which have contributed to confusion as much as to clarity (Bray 2010: 5). The first round of TIMSS, in 1995, had a question about extra lessons received before or after school; but it did not ask whether those lessons required separate payment, and thus could have included lessons in which teachers helped students free of charge. The question was repeated in 1999 and adjusted in wording in 2003 but again without asking about payment. Thereafter, the question was dropped. Thus, although the 1995, 1999 and 2003 datasets can be used to correlate extra lessons with academic achievement, they cannot be used to correlate private supplementary tutoring with academic achievement.

A similar problem has arisen with PISA (Bray 2010: 7–8). The 2000 and 2003 rounds did ask students about private tutoring, but in an ambiguous way that overlapped with other categories. Thus, the 2000 questionnaire, for example, asked about courses in test language, remedial courses in test language and private tutoring, to which respondents were invited to reply about receipt during the previous 3 years as “no, never,” “yes, sometimes” and “yes, regularly.” The question about private tutoring seemed to imply fee-payment, though it was not made explicit, and the question about courses in the test language did not ask whether or not they were fee-paying. The 2006, 2009 and 2012 rounds only asked about out-of-school-time lessons without asking whether or not they required separate payment. As a result, the PISA international datasets have also been of limited usefulness for study of the impact of private supplementary tutoring.

Large national surveys

Alongside, these large international surveys are various national ones. A useful starting point is the study in Bangladesh by Nath (2008). He analyzed 1998 data collected by a non-governmental organization from 33,229

households. The survey included questions about participation in tutoring and also tested the basic education competencies of 3,360 students. The survey found that 49.8 % of students with tutoring satisfied the basic education criteria, compared with 27.5 % of students without tutoring (p. 65). A second test of students in Grade 5 administered in 2000 found that students with tutors achieved 17.4 competencies compared with 15.5 competencies among those without tutors (p. 65). Nath reported effort to ensure the validity of the instruments and the reliability of the data, but he could not make strong causal claims that the students scored higher *because* of the tutoring.

With this consideration in mind, the study of mathematics achievement of 10,013 Grade 9 students in Taiwan by Kuan (2011) seems more useful. Kuan analyzed data gathered by the Taiwan Education Panel Study (TEPS) in 2003. TEPS had previously collected data on the same students in 2001, which therefore provided some contextual information over time. Kuan used Propensity Score Matching (PSM)² to analyze the mathematics achievements of students receiving and not receiving private tutoring and used other TEPS data from parents and teachers to control for students’ socioeconomic status, ability and attitude.

Kuan’s study shed light on the variables that shaped the students’ achievement. He found that students who received tutoring were on average more studious and from higher social classes. As one might expect, gains from tutoring were greater among students who were motivated, but in both groups the gains were small (p. 362). A major weakness of Kuan’s study was that all types of tutoring, i.e., ranging from one-to-one to large classes, were merged into a single variable which, perhaps inappropriately, was collectively labeled “cram schooling” (p. 344). The data were also limited to tutoring received during a single semester of Grade 9 (p. 353). Further, the paper made no distinctions between levels of intensity of tutoring and simply coded the tutoring as yes/no.

A related study by Liu (2012) used part of the same database but focused on 2001 when the students were in Grade 7. Whereas Kuan grouped all types of tutoring under the label of cram schooling, Liu was specifically concerned with institutions that provided examination drilling (p. 47) and excluded individual and small-group tutoring. Liu conducted a multivariate regression analysis which allowed for different hours spent in cram schools (i.e., not simply yes/no). After controlling for other variables, Liu found

¹ At the outset, in 1995, TIMSS referred to the *Third International Mathematics and Science Study*. This name was retained in 1999 when the survey was repeated. The name *Trends in International Mathematics and Science Study* has been used since 2003.

² PSM, developed Rosenbaum and Rubin (1983), is one technique to allow for selection bias in econometric analysis. Several recent studies of shadow education (e.g., Kuan 2011; Lee 2013; Thapa 2011) use PSM to enhance the validity of causal inferences in their non-experimental research.

significant positive effects of tutoring on analytical ability and mathematics performance, but the positive effects decreased when tutoring hours were lengthened. Liu also noted (p. 51) that children of highly educated parents were less likely to attend cram schools, perhaps because these parents did not approve of the institutions. It seems likely that such families would invest in one-to-one or small-group tutoring, which was excluded from analysis. Like Kuan, Liu focused only on mathematics.

A similar study was conducted by Byun (2014) in the Republic of Korea. Byun began by reviewing the mixed findings of various researchers on the Korean Youth Panel Surveys, the Korean Educational and Employment Panel and the Korean Education Longitudinal Study (KELS). Byun proceeded to his own analysis of KELS, which delivered annual data from a cohort of approximately 7,000 randomly selected students who had been in Grade 7 in 2005. KELS both provided longitudinal data on participation in tutoring and included a mathematics achievement test. KELS therefore permitted investigation of the impact of tutoring on achievement gains as measured by this test.

Byun analyzed mathematics achievement data for students moving from Grade 7 to Grade 9, combining ordinary least squares (OLS) regression with PSM and assessed separately the roles of cram schooling, individual or group tutoring, mail-based correspondence courses, internet tutoring and the government's Educational Broadcasting System courses. He found that cram schooling had a significant positive effect on achievement, but that other forms of tutoring did not. Explanations included that most cram schools closely followed the school curriculum and provided practice examinations and that many cram schools developed their own tools for curriculum and assessment. In addition, cram schools generally served high achievers who had more social capital. Nevertheless, Byun recognized limitations of his study, including that KELS did not contain panel weights and thus could not be generalized. Also, each form of tutoring received by the students was simply coded yes/no and had no indicators of quality. Further, Byun recognized (p. 56) that "although the PSM method is useful for correcting selection effects based on observed characteristics, there may be unmeasured variables related to the self-selection into the use of a particular form of tutoring, which were not considered in this study." Thus, although Byun's study was an advance on previous work, it still had limitations.

Other analysts of KELS include Lee (2013) and Ryu and Kang (2013), who have in parallel used PSM with other methods to estimate the causal effects of dimensions of tutoring on verbal, English and mathematics competencies in 2005, 2006 and 2007. Ryu and Kang (2013) reported only overall effects, but Lee reported details on each competency by year. Lee reported statistically significant

positive effects of tutoring on students' academic achievements in middle school, while Ryu and Kang reported modest effects. An explanation of the different findings from the same data and the same methods lies in the use of different aspects of tutoring as an independent variable. Lee used dichotomous variables that indicated (a) total years of private tutoring and (b) private tutoring participation in general (p. 66), while Ryu and Kang used expenditure on tutoring as an independent variable. Thus, findings can greatly differ even when using the same data and methods. Moreover, neither of these studies made full use of the differentiation within the KELS data between tutoring in *hagwons* (cram schools), by individual tutors, via the internet and through broadcasting media.

In a different context, Dang (2007) analyzed 1997/1998 national household survey data in Vietnam using a joint Tobit and ordered probit econometric model. The survey collected data on tutoring expenditures and on students' academic performance in the previous grade. Data on academic performance were collected from the students themselves or from other household members, with the four values of excellent, good, average or poor. Dang found positive correlations between tutoring expenditures and achievement, noting (p. 696) a particularly strong impact at lower secondary compared to primary schooling, except for the pupils who were low performers. Again, however, the dataset had limitations. The questionnaire only asked for expenditures on tutoring without specifying the type and content of classes. Thus, the survey omitted indicators of quality and curriculum, and the responses could have included non-academic subjects. Also, the students' self-reported data on their attainment as excellent, good, average or poor may not have been completely trustworthy.

Perhaps for these and other reasons, Dang's findings only partly matched those of Ha and Harpham (2005), who analyzed data from 1,000 8-year-old children randomly selected from 4,716 households in 2002. The focus of extra classes included dancing, swimming, singing, chess and painting as well as mathematics and Vietnamese language. The test of academic achievement was very simple, i.e., ability to write a simple sentence and multiply two by four. After controlling for ethnicity, household wealth, region and other factors, the researchers found that receiving extra classes was not significantly associated with the children's writing and numeracy.

Smaller-scale investigations

While large datasets are often impressive, valuable insights can also be gained from smaller studies. Beginning again with Bangladesh, Hamid et al. (2009) surveyed 228 Grade 10 students in eight rural schools and interviewed 14

pupils. The study focused on the learning of English, and instruments included a student survey questionnaire, an English proficiency test, school records of students' grades and results of the public Secondary School Certificate (SSC) examination. In the quantitative analysis, the researchers found through ordered logistic regression that students who had received private lessons had double the frequency of higher grades than their counterparts who did not receive private lessons and that the only other variables significantly associated with achievement were gender and mother's education (p. 293). However, this finding again was based on a dichotomous yes/no coding of private tutoring. Further, each model represented approximately 10 % of the total variability, which limited the validity of each analytical model.

In this light, the interviews were especially helpful. Three of the 14 students had never received private lessons, and four had received tutoring in the past but had stopped. One had had a home tutor since Grade 4 and had received tutoring every day, while another had only commenced receipt of tutoring a month previously. The interviewees repeatedly emphasized the poor quality of English teaching at school, complaining that classes were irregular and that teachers were incompetent and/or uncommitted (p. 299). This may indeed reflect patterns in Bangladesh which has much inferior schooling to that in the Republic of Korea, for example. At the same time, the researchers highlighted a social chain effect (p. 302) in which interviewees referred to peers who were receiving tutoring and who led them to believe that tutoring was necessary for enhanced learning. It appeared that none of the 14 students was being tutored by their regular classroom teachers, though this practice is common in Bangladesh and has been considered a form of corruption (Manzoor 2013: 20; Nath 2008: 66). The researchers recognized the limitations of their work, including the need for more focused and discriminating categories of analysis to investigate the types, quantities, intensities and qualities of tutoring.

Further along the spectrum of qualitative research is a study in Malta by Gauci and Wetz (2009). The researchers examined styles and content of teaching received by 18 pupils in a Grade 11 mathematics classroom, 16 of whom were receiving various forms of private supplementary tutoring. The researchers observed school classes twice a week for approximately 2 months (p. 38) and interviewed in depth not only the selected pupils but also their teacher. They analyzed the experiences of 12 pupils including the two who did not receive supplementary tutoring. These students all had different characteristics. Two were lower achievers taking a different examination that catered for that level, Paper IIB, who felt neglected because the teacher devoted most lessons to the majority who were taking the more demanding Paper IIA. The teacher was aware that

both students were receiving extra tutoring and felt that in for one it could be a good investment but that the other had neglected her studies for too long for tutoring to be able to salvage the situation. In the event, neither of them passed; but Gauci and Wetz (2009: 68) seemed convincing when stating that in the circumstances of relative neglect by the teacher, this pair of students "could ill-afford not to attend private tuition that was specifically tailored for them."

Concerning the two students who did not receive tutoring, Gauci and Wetz recorded that one scored Grade 1 (i.e., the top) through her own efforts, but that the other scored only Grade 3. The pattern among the other eight students who received tutoring was mixed. One gained a (passing) Grade 5, and "it was probably useful for her to attend private lessons as it allowed her to practice more and gave her a second chance to understand things that she did not understand at school" (p. 69). Two other students failed Paper IIA, and indeed should probably have been guided to take the less demanding Paper IIB. For them, "private lessons seemed not to work, but this could have resulted from their 'wrong' choice of paper" (p. 69). The other five students achieved results in the middle ground for which it was difficult to speculate about the effect of the private lessons. The researchers were appropriately cautious about the extent to which their qualitative study of 12 students could be generalized to the whole education system, but their work nevertheless provided many insights.

Also worth mentioning is a rare quasi-experimental study conducted by Mischo and Haag (2002) in Germany. The researchers provided tutoring for 122 pupils in Grades 5–11 and asked them to find schoolmates with approximately the same combination of subject matter and performance who were not receiving tutoring. Students in the first group received tutoring for 90 min a day, 4 days a week, in clusters of four pupils homogeneous with respect to age and subject matter. The tutors were recruited by tutorial centers with registered offices throughout Germany. Baseline data were collected 1 month after term had commenced, with assessments of text anxiety, self-concept of ability, action control and learning motivation, and with school marks in mathematics, Latin, English and French. The comparison data were collected at the end of the school year, 9 months later.

The researchers found that pupils receiving paid tutoring received significantly higher school marks than their counterparts without tutoring (Mischo and Haag 2002: 270). The pupils with tutoring also showed improvement in motivational variables. However, no differences were observed in action control, perhaps because tutoring for 4 days a week for 90 min each did not require self-regulating strategies and therefore action control. The researchers added that whether improvement in school marks was the consequence or cause of improved

motivational variables could not be clearly answered “since emotional development and school achievement not only mutually affect each other, but also affect themselves individually within the learning process” (p. 270). The researchers added that the detailed processes of tutoring and their causal impact on motivational variables and learners’ use of cognitive and metacognitive strategies could not be investigated in their study.

Conclusions

This paper began by noting first the global expansion and significance of shadow education and second the question for stakeholders whether shadow education “works” in the sense of raising academic achievement. The introduction observed confident positive assertions not only from the tutoring sector itself but also from journalism and academia. However, it added that careful analysts are ambivalent. As remarked for instance by Byun (2014: 40), “empirical evidence has been inconsistent, contradictory and confusing.” The examples from the empirical literature cited in this paper echo Byun’s observation. In line with the title of the paper, this concluding section first summarizes why the research is inconclusive and then indicates what can be done about it.

Why the research is inconclusive

The first problem is that foci are defined imprecisely. Shadow education can mean different things to different people. The present paper, in line with authors such as Bregvadze (2012), Lee et al. (2009) and Silova (2009), is concerned with tutoring in academic subjects provided in exchange for a fee. However, the TIMSS and PISA questions embraced fee-free as well as fee-paying tutoring and could include extra lessons provided by teachers as part of their regular work. The Vietnamese data reported by Dang (2007) and Ha and Harpham (2005) focused on fee-paying instruction but included non-academic as well as academic activities. A similar remark applies to the research in China by Xue and Ding (2009).

A second problem of definition is in the nature of tutoring, which embraces a much broader range of modes than regular schooling. Kuan (2011) merged all forms of extra academic instruction in TEPS into a single category labeled cram schooling. Yet while some tutoring has a cramming function focused on intensive preparation for examinations, other tutoring provides enrichment and broadening. Further, the large classes for which cram schooling is well known are very different from one-to-one and small-group instruction. Liu (2012), who also analyzed TEPS data, explicitly excluded individual and small-group

tutoring from the analysis. The fact that Kuan and Liu both referred to cram schooling in the titles of their articles contributed to confusion. Only careful readers will realize that one has a wider lens than the other even though they both use TEPS data and purport to analyze cram schooling.

Byun’s analysis of KELS avoided this problem by separate consideration of cram schools, private (one-to-one or small group) tutoring, mail correspondence courses, internet tutoring and educational broadcasting. However, Byun coded the students’ receipt of the different types of tutoring as yes/no and did not examine either different durations or different qualities of tutoring. In this respect, Byun’s work had the same weakness as that of Kuan (2011). Both Byun and Kuan were persuasive in stressing the merits of PSM in contrast to OLS regression. However, the process of matching can only use observed characteristics and omits unobserved variables that could be important. The findings of PSM can be further justified or confirmed by sensitivity analysis to check whether omitted variables were important enough to change (or reject) the findings from PSM (Austin 2011).

The question about the qualities of tutoring raises further issues because qualities are difficult to define and measure (Acedo et al. 2012; UNESCO 2005). Thus, it cannot be assumed that one-to-one tutoring has higher quality than small-group or large-class tutoring—much depends on the quality of the tutor, and indeed one reason why classes are large may be that students perceive the quality to be high and therefore decide to enroll. Moreover, some students find that large classes with didactic teaching fit their learning styles better than small classes with interactive teaching (Yung 2011; Zhang et al. 2012).

Quantitative researchers might argue that from an empirical perspective, it does not greatly matter how students gain higher grades if the research question simply seeks a correlation (or preferably causation) between tutoring and learning. However, this raises further questions how learning is measured and to what extent test scores indeed measure cognitive achievement. Mischo and Haag (2002: 267) recognized that school marks in the four subjects of their study might be problematic as indicators of cognitive achievement, but stated that “construction of achievement tests in each subject matter ... would have been beyond the capabilities of the present study.” Hamid et al. (2009) did crosscheck their test grades in English with the scores achieved by students in the public SSC examination. They found particular variation at the extremes—but could not decisively say which test provided a better measure of achievement or, indeed, whether either of them was fully trustworthy. Byun merely reported that his measure of achievement had been devised by the Korean Educational Development Institute using Item Response Theory scores and did not comment on the

reliability of this test. Similar remarks apply to the tests in the TEPS data analyzed by Kuan (2011) and Liu (2012).

Finally, none of the studies discussed seemed to recognize that private tutoring may subtract as well as supplement. The most obvious subtractive situation arises when teachers deliberately reduce their effort in regular hours in order to promote the market for out-of-school private classes with the pupils for whom they are already responsible. Among the countries on which this paper has focused, the practice is common in Vietnam and Bangladesh (Dang 2011: 26; Nath 2008: 66). In countries where teachers are prohibited from tutoring their own students, they may nevertheless tutor other students and put more effort into this private-sector activity than into their regular duties. And in countries where teachers rarely give any private tutoring, they may assume that support is readily available in the marketplace and therefore not put so much effort into their own duties.

What can be done about it

Given these remarks, the final question concerns the research agenda ahead and the strategies to accomplish it. First, the research question should be refined. The question “Does private supplementary tutoring work?” is too broad to be meaningful. It needs to be rephrased along the lines of “What types, qualities and quantities of private supplementary tutoring, with what durations, intensities and backup support, work in what types of learning domains for what sorts of students in what sorts of circumstances?” This complex question fits better the complex reality. Studies which aggregate types of tutoring and which fail to take account of the quantities and qualities of tutoring are arguably uninformative and perhaps misleading. Presented with the authority of precise numbers and complex formulae, they may appear scientific but in practice lead to conclusions of limited practical value.

To some extent, this paper has traversed familiar ground by contrasting quantitative, qualitative and mixed approaches to research. As noted by many authors (e.g., Fairbrother 2014; Johnson and Christensen 2012), each approach has both merits and limitations. Quantitative studies, especially those with large random samples, harness multiple variables and have the potential to demonstrate mathematical relationships. If carefully designed and executed, panel data such as TEPS and KELS can go further with insights into causation rather than correlation. However, qualitative studies provide insights that cannot be secured through the quantitative approaches. Gauci and Wetz (2009) demonstrated the personal circumstances of individual students in the specific environment of a class taught by a teacher with her own characteristics and biases. The nuanced picture indicates why the long and complex

research question is in practice more meaningful than the short and superficial one. Hamid et al. (2009) adopted a mixed-methods approach in which the qualitative data exposed and even contradicted some aspects of the picture from the quantitative approach.

Also among the studies on which this paper has focused has been the quasi-experimental approach of Mischo and Haag (2002). Byun (2014: 56) concluded his paper with a call for more experimental work. Certainly, it would be desirable from a conceptual perspective, though ethical issues may arise when educational inputs are given to some children and deliberately withheld from others. The study by Mischo and Haag was small and arguably did not disadvantage the control group since these students evidently did not plan to seek tutoring; but larger and more demanding studies might encounter different issues.

In the absence of experimental work, stronger data on causal factors rather than correlations can be gained from longitudinal work. TEPS provides a model, though it appears that for specific data on the effects of private tutoring, the data from the initial pair of cohorts were insufficiently precise and comprehensive to permit clear identification of patterns (Kuan 2011; Liu 2012). This is partly because the surveys had other objectives and private tutoring was not a priority item. Similarly, the Vietnamese household survey utilized by Dang (2007) had many foci among which tutoring was relatively minor. Thus, one way forward is to design more studies which specifically aim to identify the impact of different types and amounts of tutoring on different types of learner in different types of context. Other objectives may be added to the surveys, but if tutoring is placed in the foreground it is likely to have more precise questions that will lead to more precise answers.

Finally, models for analysis need to note the interaction between out-of-school tutoring and in-school lessons, taking more account of the backwash of tutoring and the fact that it may subtract as well as supplement. This also is best done through qualitative studies since questionnaire surveys cannot easily secure data on power dynamics including corruption and the condoning of inefficiencies in education systems.

References

- Acedo, C., Adams, D., & Popa, S. (Eds.). (2012). *Quality and qualities: Tensions in education reforms*. Rotterdam: Sense.
- Aslam, M., & Atherton, P. (2012). *The ‘shadow’ education sector in India and Pakistan: The determinants, benefits and equity effects of private tutoring*. ESP Working Paper 38, Budapest: Education Support Programme, Open Society Foundations.
- Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46, 399–424.

- Baily, S. (2012). Framing the World Bank education strategy 2020 to the Indian context: Alignments, challenges, and opportunities. In C. S. Collins & A. W. Wiseman (Eds.), *Education strategy in the developing world: Revisiting the World Bank's education policy* (pp. 371–393). Bingley: Emerald Books.
- Baker, D. P., Akiba, M., LeTendre, G. K., & Wiseman, A. W. (2001). Worldwide shadow education: Outside-school learning, institutional quality of schooling, and cross-national mathematics achievement. *Educational Evaluation and Policy Analysis*, 23(1), 1–17.
- Bray, M. (1999). *The shadow education system: Private tutoring and its implications for planners*. Paris: UNESCO International Institute for Educational Planning (IIEP).
- Bray, M. (2009). *Confronting the shadow education system: What government policies for what private tutoring?*. Paris: UNESCO International Institute for Educational Planning (IIEP).
- Bray, M. (2010). Researching shadow education: Challenges and directions. *Asia Pacific Education Review*, 11(1), 3–13.
- Bregvadze, T. (2012). Analysing the shadows: Private tutoring as a descriptor of the education system in Georgia. *International Education Studies*, 5(6), 80–89.
- Buchmann, C. (2002). Getting ahead in Kenya: Social capital, shadow education, and achievement. In B. Fuller & E. Hannum (Eds.), *Schooling and social capital in diverse cultures* (pp. 133–159). Amsterdam: JAI Press.
- Byun, S. Y. (2014). Shadow education and academic success in South Korea. In H. Park & K. Kim (Eds.), *Korean education in changing economic and demographic contexts* (pp. 39–58). Dordrecht: Springer.
- Coniam, D. (2013). Private tutorial schools in Hong Kong: An examination of the perceptions of public examination re-takers. *Asia Pacific Education Research*. doi:10.1007/s40299-013-0113-x.
- Crotty, J. M. (2012). Global private tutoring market will surpass \$102.8 billion by 2018. *Forbes Magazine*. Accessed November 30, 2013, from <http://www.forbes.com/sites/jamesmarshallcrotty/2012/10/30/global-private-tutoring-market-will-surpass-102-billion-by-2018>.
- Dang, H. A. (2007). The determinants and impact of private tutoring classes in Vietnam. *Economics of Education Review*, 26(6), 648–699.
- Dang, H. A. (2011). A bird's-eye view of the private tutoring phenomenon in Vietnam. *IIAS Newsletter*, 56, 26–27.
- Dawson, W. (2010). Private tutoring and mass schooling in East Asia: Reflections of inequality in Japan, South Korea, and Cambodia. *Asia Pacific Education Review*, 11(1), 14–24.
- Fairbrother, G. (2014). Quantitative and qualitative approaches to comparative education. In M. Bray, B. Adamson, & M. Mason (Eds.), *Comparative education research: Approaches and methods* (2nd ed., pp. 71–93). Hong Kong: Dordrecht: Comparative Education Research Centre, The University of Hong Kong; Springer.
- Gauci, D., & Wetz, S. (2009). *The private lessons phenomenon in a form five girls' mathematics classroom*. B.Ed. (Hons.) dissertation, University of Malta.
- Growing Stars. (2013). *Growing stars: Online tutoring at its best*. Accessed September 1, 2013, from <http://www.growingstars.com>.
- Ha, T. T., & Harpham, T. (2005). Primary education in Vietnam: Extra classes and outcomes. *International Education Journal*, 6(5), 626–634.
- Hamid, M. O., Sussex, R., & Khan, A. (2009). Private tutoring in English for secondary school students in Bangladesh. *TESOL Quarterly*, 43(2), 281–308.
- Johnson, B., & Christensen, L. (2012). *Educational research: Quantitative, qualitative, and mixed approaches*. Los Angeles: SAGE.
- Kuan, P. Y. (2011). Effects of cram schooling on mathematics performance: Evidence from junior high students in Taiwan. *Comparative Education Review*, 55(3), 342–368.
- Kwo, O., & Bray, M. (2011). Facing the shadow education system in Hong Kong. *IIAS Newsletter*, 56, 20.
- Lee, J. Y. (2013). *Private tutoring and its impact on students' academic achievement, formal schooling and educational inequality in Korea*. PhD dissertation, Columbia University.
- Lee, C. J., Lee, H., & Jang, H. M. (2010). The history of policy responses to shadow education in South Korea: Implications for the next cycle of policy responses. *Asia Pacific Education Review*, 11(1), 97–108.
- Lee, C. J., Park, H. J., & Lee, H. (2009). Shadow education system. In G. Sykes, B. Schneider, & D. N. Plank (Eds.), *Handbook of education policy research* (pp. 901–919). New York: Routledge for the American Educational Research Association.
- Liu, J. (2012). Does cram schooling matter? Who goes to cram schools? Evidence from Taiwan. *International Journal of Educational Development*, 32(1), 46–52.
- Manzoor, S. (2013). *Reasons and necessity of private tutoring in English for Bangla medium primary school students in Bangladesh*. Thesis, Department of English and Humanities, BRAC University.
- Medway, F. J. (1995). Tutoring. In L. W. Anderson. (Ed.), *International encyclopedia of teaching and teacher education* (pp. 271–274). Oxford: Pergamon Press.
- Mischo, C., & Haag, L. (2002). Expansion and effectiveness of private tutoring. *European Journal of Psychology of Education*, XVII(3), 263–273.
- Mori, I., & Baker, D. (2010). The origin of universal shadow education: What the supplemental education phenomenon tells us about the postmodern institution of education. *Asia Pacific Education Review*, 11(1), 36–48.
- Nath, S. R. (2008). Private supplementary tutoring among primary students in Bangladesh. *Educational Studies*, 34(1), 55–72.
- Rohlen, T. P. (1980). The *juku* phenomenon: An exploratory essay. *Journal of Japanese Studies*, 6(2), 207–242.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.
- Ryu, D., & Kang, C. (2013). Do private tutoring expenditures raise academic performance? Evidence from middle school students in South Korea. *Asian Economic Journal*, 27(1), 59–83.
- Silova, I. (Ed.). (2009). *Private supplementary tutoring in Central Asia: New opportunities and burdens*. Paris: UNESCO International Institute for Educational Planning (IIEP).
- Silova, I., Būdienė, V., & Bray, M. (Eds.). (2006). *Education in a hidden marketplace: Monitoring of private tutoring*. New York: Open Society Institute.
- Stevenson, D. L., & Baker, D. P. (1992). Shadow education and allocation in formal schooling: Transition to university in Japan. *American Journal of Sociology*, 97(6), 1639–1657.
- Thapa, A. (2011). *Does private school competition improve public school performance? The case of Nepal*. PhD dissertation, Columbia University.
- Tutors International. (2013). *About tutors international*. Accessed September 21, 2013, from <http://www.tutors-international.com/about>.
- Ünal, H., Özkan, M., Milton, S., Price, K., & Curva, F. (2010). The effect of private tutoring on performance in mathematics in Turkey: A comparison across occupational types. *Procedia Social and Behavioral Sciences*, 2(2), 5512–5517.

- UNESCO. (2005). *Education for all—The quality imperative. EFA global monitoring report 2005*. Paris: UNESCO.
- Ventura, A., & Jang, S. (2010). Private tutoring through the internet: Globalization and offshoring. *Asia Pacific Education Review*, 11(1), 59–68.
- Xue, H. P., & Ding, X. H. (2009). A study on additional instruction for students in cities and towns in China. *Educational Research*, 30(1), 39–46 (in Chinese).
- Yung, W. H. (2011). *Shadow education in Hong Kong: The experience of learners of English*. MA in applied linguistics dissertation, The University of Hong Kong.
- Zeng, K. (1999). *Dragon gate: Competitive examinations and their consequences*. London: Cassell.
- Zhang, L. F., Sternberg, R. J., & Rayner, S. (Eds.). (2012). *Handbook of intellectual styles: Preferences in cognition, learning, and thinking*. New York: Springer.