

How can we unfold the reality of student learning, if there is a reality: Pitfalls – and bridges – of educational research

Wee Tiong Seah¹ and Ngai-Ying Wong²

¹The University of Melbourne, Graduate School of Education, Australia (ORCID: 0000-0001-9378-8439)

²The Education University of Hong Kong, Department of Curriculum and Instruction, Hong Kong (ORCID: 0000-0002-9502-662X)

Article Info	Abstract
<p>Article History Submitted: 30 October 2018 Revised: 15 March 2019 Published online: 1 April 2019</p> <hr/> <p>Keywords Educational research Research methodology Mixed methods Data analysis Student learning</p>	<p>This paper discusses some of the recurrent issues which the authors have noticed in educational research. These might be concerned with the nature of social science research, such as its representation of the reality of student learning, the role for replication studies, and the tolerance for disagreements. There are also issues related to the researching process, including the scope of prior research that should be reviewed, the purpose of triangulation, the need to be data-sensitive, the value in the complementarity of quantitative and qualitative information, relationships between theory and data, researchers' statistical literacy levels, and over-dependence on software-generated results. The fluid multi-cultural context in which education and pedagogy are situated has also meant that accessibility to knowledge is language-mediated and social reality, ever-evolving.</p>

"Say what you know, do what you must, come what may."

Sofia Kovalevskaya (1850-1891)

1. Prologue: Learning by doing and reflecting

For many academics, the natural professional journey might be one of 'imitating others' research during one's own graduate studies – developing one's own research agendas and conducting them both independently and collaboratively – supervising one's research students and reviewing others' manuscripts – leading a large funded project'. Some might even have the additional experience as part of their post-retirement consultancy work of helping academics to submit their research proposals and journal articles, like what the second author is doing for the time being. At the start, it is quite common that one is quite ignorant of what educational research is all about. Some academics were initially trained in a non-education related field too. Both authors, for example, were qualified as mathematicians first in their respective countries. Gradually one begins to establish his/her research profile in a certain area (e.g. mathematics education). As one proceeds along the academic career path, one is also expected to take up thesis supervision, paper/project proposal reviews and the role of an external examiner in a wide range of fields, from, say, mathematics education to curriculum studies, educational administration, educational psychology

Address of Corresponding Author

Wee Tiong Seah, Melbourne Graduate School of Education, The University of Melbourne 234, Queensberry Street, The University of Melbourne, Victoria 3010, Australia.

✉ wt.seah@unimelb.edu.au

How to cite: Seah, W.T. & Wong, N-Y. (2019). How can we unfold the reality of student learning, if there is a reality: Pitfalls – and bridges – of educational research. *Journal of Pedagogical Research*, 3(1), 92-99. doi: [10.33902/JPR.2019.7](https://doi.org/10.33902/JPR.2019.7)

and even psychology in general. One often socializes oneself into a field by doing, as well as through collegial exchange. Reflection is of utmost importance too. Having travelled on this journey to different extent, the authors have become aware of some recurrent issues that are commonly experienced among researchers, regardless of whether they are novices or experienced. We think it worthwhile to share some of these recurrent issues here, hoping that the discussion could stimulate further discussions, and be of help to the educational research community.

2. The blind man and the elephant: The pattern of scientific discovery

There is an ongoing debate whether, unlike scientific research, one can really 'discover' the 'truth' or unfold the reality in social science research, including educational research. The essence of science lies in replicability. Some emphasizes generalizability in social science studies. Yet, more often than not, results in educational research are either contradictory or inconclusive. Even if we consider so-called well-established results like how affect (self-confidence, self-concept, motivation, etc.) affects learning outcomes, we must admit that generally the accountable percentage is not large, not to mention the existence of contradictory results. Tan and Li (2015), for example, provided examples of research studies which reported contradictory findings of how positive and negative affect can both be related to creativity performance. Other examples include Kagan (1992) and Ma (1999). The same is true for research into gender differences, as noted by Leder (1992) more than 25 years ago, and also by Vale and Bartholomew (2008) more recently. We also see in the review of Dowker et al. (2016) how there exists contradiction in research findings connecting gender and mathematics anxiety. It reminds us of the Indian parable of 'the blind men and an elephant'. When we conduct an educational research, each one of us might see ourselves as contributing a dot on the paper. As more of us do the same, we come up with a picture of what we collectively have been trying to portray. There is room for much misunderstanding and wrong guesses initially. For example, the dots that we have initially might lead us to think that the emerging picture is that of a snake. We would need more dots spreaded out across a broader area of the eventual picture - representing more of us contributing more findings using a variety of approaches and perspectives - before we develop a better and more accurate picture that is the elephant, and recognizing that the 'snake' is in fact the trunk of the elephant (Figure 1)!

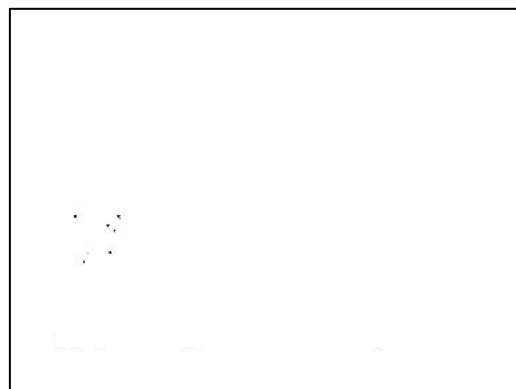


Figure 1(a). Some dots to start with

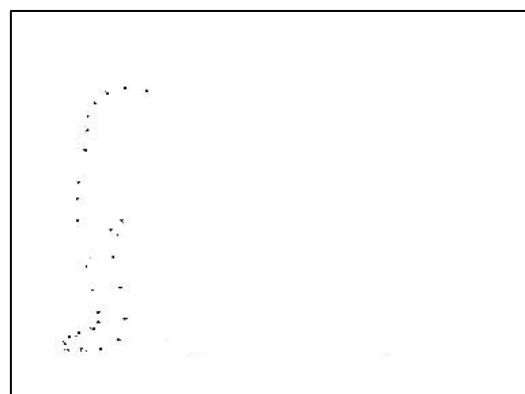


Figure 1(b). It's a snake!

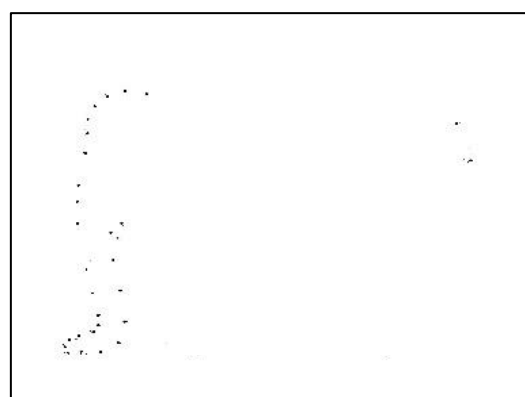


Figure 1(c). Just a few irregularities

In a sense, there is not much difference with scientific research (mathematics research could be an exception – it basically involves logical deduction from a set of postulates, however mathematics does have its realistic and cultural origins which we won't run into the discussion here, see Wilder, 1952). One of the first such revisions of previously-accepted models is Hooke's law, which in simple terms stipulates that the elongation of a spring is proportional to the force applied to it, and which is the principle behind such tools as the spring scale. To begin with, the law did not come from theoretical but experimental considerations. The British physicist Robert Hooke (1635-1703) saw a pattern and came up with a law. It was revised when more and more 'outlier' results were observed. We have more examples in optics and even in Newtonian mechanics. So we need an abundance of studies to help us grasp the complete picture. In a sense, 'truths' are not established once and for all. Like the 'join-the-dots game', we need enough dots before a more complete picture emerges. Yet, how do we know that we have done enough to know that what looks like the 'snake' before us is not a valid picture of the phenomenon or situation?

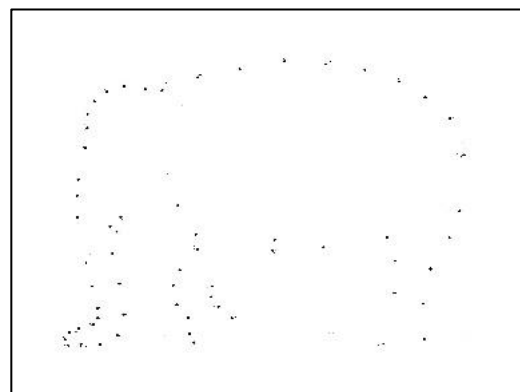


Figure 1(d). An elephant!

3. Where do frameworks come from?

We need an abundance of studies (though they do keep us happily busy, and employed!) and the course of uncovering the picture could be long. It may take generations. But we don't do it blindly. We have to make sense. That's why it is important for us to understand the discourse that is concerned with what our forerunners have done, how far they had travelled, what direction the route in general is heading towards, what challenges they had met, what have been the irregular results, and what the opportunities for sharpening methodologies were. Indeed, this is precisely the main purpose of conducting literature reviews, not just quoting a few famous and/or convenient publications and then structuring a study based on these. What we need to guide the design of any research study is an extensive literature review. This by no means refers to a long list necessarily, but rather, it should piece up existing literature representing a variety of approaches and contexts for it to showcase the development trend of the research topic, making up what might be called the 'talk of the town'.

With a reasonably rich mapping of what had been researched before, one begins to conceptualize what has been happening so far in the context of the research topic. This conceptualization is essentially what we might refer to as 'framework' (at least an operationalized one). For example, the second author (with his collaborators) conceptualized the notion of *lived space* (Figure 2) (Wong, Marton, Wong, & Lam, 2002), which would be gradually revised and extended in time to accommodate more research agendas (e.g. teachers' knowledge and belief, in Zhang & Wong, 2015, as shown in Figure 3; and even religious worldview, in Leu, Chan, & Wong, 2015).

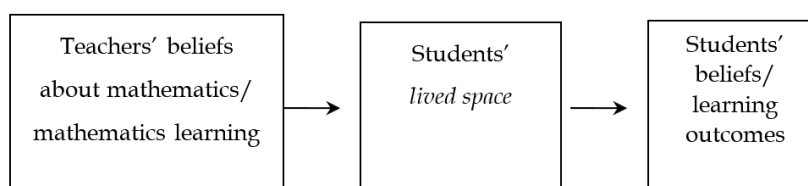


Figure 2. The *lived space* of mathematics learning

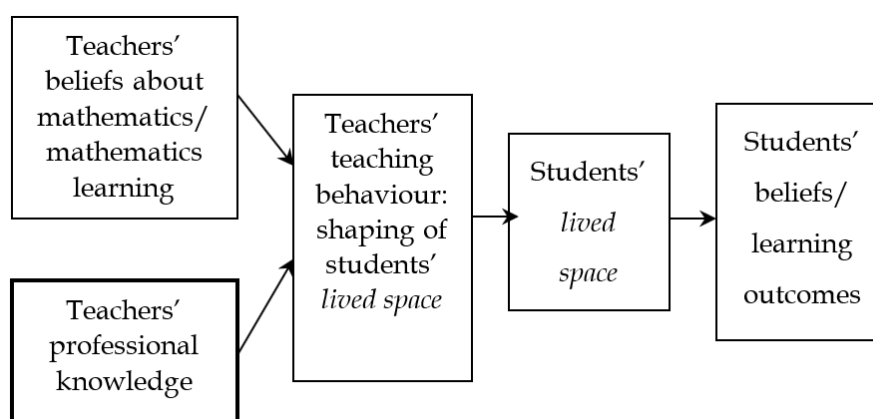


Figure 3. Teachers' professional knowledge, beliefs and their teaching

To grasp the trend of intellectual discourse (i.e. the so-called 'talk-of-the-town' mentioned above), the scope of the literature reviewed plays an important role. It must be extensive. The problem is that educational research is disseminated in different cultures in different languages. Though a large portion of research reports nowadays are published in English, there are also established and active research traditions and activities (including the different modes of dissemination) around the world where the language of communication is not English, such as in China, Germany, Japan and Korea. In this sense, monolingual researchers or research teams might embark on research designs without access to – and knowledge of – what other cultures might already know in the respective research fields.

For those more 'regional' research students, there is always a risk of confining their literature to their own regional language. It is even more worrying if one relies solely on translations (which may not be that accurate – and there are also cultural differences in the use of terminologies) and second-hand resources. For example, a request by a teacher for a student's parents 'to meet him/her' may reflect a collaborative parent-teacher relationship in some cultures, yet it may also imply tension in another one. Certainly, there is much to gain if bilingual researchers review research publications in different languages, and not regarding the English-speaking literature as 'international'.

4. How much worth are replication studies?

Prof. Mogens Niss had in a 2018 conference keynote address commended a leading mathematics education research journal for promoting replication studies, that is, studies in which published research are (re-)analysed and discussed (Niss, 2018). Just like the establishment in medical science of the effectiveness of drugs, an abundance of studies is required to establish a 'fact' in social science (including the establishment of the validity of an instrument). In fact, in Chinese medicinal practices, famous doctors in different times had even published books on treatment cases, which serve as important reference books for the later generations. Thus, replication studies have their own worth, as long as one knows precisely what one is aiming for in doing so, whether it is for refining the data-collection instrument, refining the framework, or clarifying outlier results in previous studies, at the same time informing the academic community what lessons can be learnt in doing so. Mere (and blind) replications do not have much worth.

Now, there are times when we respond to what look like outlier results by adding potential factors to see if a certain phenomenon can be better explained. We may 'apply' previous studies to new target groups or we may investigate an issue again under a new situation. An example of the latter was when the learning environment studies were 're-investigated' again under the constructivist's classroom (see Taylor, Fraser, & Fisher, 1997). In a sense, these are not mere replications, but rather, they are attempts to extend the existing academic discourse. They are

advancements. In fact, we might also argue that it is almost impossible to conduct a replication study, since it is virtually impossible to maintain the same constants and variables over space and time. Even if the same methodology was to be carried out with all the Year 7 students in a particular school 5 years apart, we cannot say with confidence that the target groups of participants are similar and comparable.

5. How mixed are mixed methods?

Mixed method has become a fashion. It seems that it is a selling point to entertain or please reviewers from both quantitative and qualitative traditions, and to claim that one knows both methods. The question is, are the methods really mixed, that is, incorporated such that one method is adding value to another? Let's look at the issue from the 'blind men and an elephant' perspective. Regardless of quantitative or qualitative methods, we need lots of 'dots' to help us compose a picture that is as accurate as possible. Thus, each method in your mixed method study – whether it is a questionnaire, interview, class observation, or journal, for examples – may also be considered to be contributing a dot to the picture you are trying to paint. Indeed, we feel that this is the real meaning of triangulation. In other words, triangulation should not involve just cross-validation using multiple methods, but it is about contributing different components to your study to help create the picture.

The notion of triangulation originated from surveying. In social science, conventionally, it refers to the use of different methods to validate the same result (including the so-called 'multi-triad-multi-method'). Gradually the notion took on a broader sense. A typical example is, when we evaluate a curriculum change (in a school), we might survey or interview students, teachers, curriculum leaders (in the school), and administrators. We might also conduct class observations. Obviously, this variety of data sources would not lead us to the same result since their contexts and perspectives are different. It is also possible that the results are conflicting: e.g. the administrators favored the change, the teachers reluctantly implemented it but 'miraculously' the students loved it (or conversely the teachers implemented with passion but students disliked the change). Nevertheless, this disagreement is the result! Any attempt to 'massage' the data (sources) to showcase a sense of unified message is simply rendering the findings invalid, and academically unethical.

6. Proving that 'mothers are women': The framework birdcage

Reviewers often ask for frameworks. But where does a framework come from? Of course it can come from 'theory' and literature. So, for example, we might borrow the learning framework of a famous scholar. But we can continue to ask, how did that scholar 'produce' her or his framework in the first place? Has her/his framework become the benchmark framework because of her/his fame? Certainly some researchers borrow overarching theories from other disciplines (e.g. learning environment from organizational theory), but the point remains that most (if not all) existing frameworks are sort of conceptualizations from previous studies. They are not cast in stone and are subject to projection (to particular research context), revision and even over-turning. Looking for a relationship among factors (that appear in a framework) under that framework is like what Hong Kong people would call it as validating that mothers are women – it's a 'tautology'. Methodologically, it is possible that a rigid framework could block 'unexpected' results that are of utmost importance to scientific advancement. For example, if we do not include parents' social economic status in the research framework of student learning in the first place, most probably we cannot identify its influence though this influence exists.

That's why exploratory studies, when purposefully-designed, have an important role to play in our quest to better understand how to promote student learning. Due to the exploratory nature of these research studies, unexpected results often produce the driving force to theoretical advancement. An example here would be the establishment of the 'big-fish-small-pond' effect, which was stimulated by the seemingly contradictory observation of low self-concepts among

gifted students in exploratory studies. Further investigations had then revealed that since these students were put together in prestigious schools, being now surrounded by peers who were also gifted had led to a sense of 'feeling small', leading to a drop in self-concept, and accounting for the 'big-fish-small-pond' effect (see Marsh & Parker, 1984).

It is important that we stay data sensitive so that we can unfold new horizons instead of getting frustrated whenever an expected (statistical) correlation does not come about. There is much room for this in qualitative methods too. It is quite common that participants would disclose something unexpected during interviews. Ignoring this unexpected information and regarding it as 'just' some noise from a single case cannot lead us to seeing the elephant while we keep seeing the snake!

7. Possessing a quantitative as well as qualitative mind

We are not sure when a line began to be drawn between quantitative and qualitative methods and these were subsequently called 'traditions'. The reality, if there is one, is something represented by an integrated whole. Given that we pointed out above that both 'quantitative' and 'qualitative' (so-to-speak) dots contribute to the complete picture, both ways of looking into a research problem are useful. For instance, sometimes it helps to frame a study in terms of dependent versus independent variables, or of causal relationships, etc. even when one is employing qualitative methods. The notion of control factors is also essential. Likewise, one needs to remain data sensitive and interact with one's data even when one is analyzing a quantitative set of data such as in questionnaires.

8. Data driven, data polishing and data massage

Most educational research cannot avoid being data driven, regardless of whether the data is qualitative or quantitative. Otherwise, we would not have used the elephant as a metaphor. Testing which factors cause which other factors (so-called fixing dependent and independent variables) arbitrarily or just choosing a final 'model' yielding the best psychometric properties not only involves ethical issues but might also indicate that one has lost one's direction. Are we then suggesting that research must be theory-driven instead of data-driven? As we mentioned earlier in this paper, 'theory' is no other than the trend of previous studies (discourse in literature). In other words, both theory and the data collected should be taken into account.

We would like to distinguish between what we mentioned above and 'data massaging', which is normally understood as the purposeful selection of what constitute data in a research study. This may take the form of freely re-grouping items (by exploratory factor analysis, say), or deletion of items (or even a whole subscale) just to improve psychometric properties. We do not support data massaging, as we believe that all information is useful and should be treated as data as such, even if the analysed data do not present a 'neat-and -tidy' picture. In fact, does this not reflect the 'messy' nature of societal phenomena, and is it not the researchers' responsibility to represent these phenomena as such, which would involve interpreting analysed data against real-world context?

One related issue is whether one can delete items (of questionnaires) after data collection. It is quite common practice to do this to improve psychometric properties, the Cronbach alphas in particular. Certainly instruments can be revised. One can imagine the analogy of medical science (though we are not learned in this field): if one tests a drug with a patient and find it not effective, it is natural that one adjusts the dosage. But isn't it true that one needs to test it again before calling it effective? Arbitrarily deleting items would fall into the trap of massaging data too (since one could come up with different psychometric properties with another set of data!). Since usually we claim that we have chosen a well-established instrument (if it is not well established, why choose it?) right from the start, we should refrain from revising it after data collection without good reasons. If poor psychometric properties come about, one should rather look for contributing reasons. The second author had an experience of re-designing a new instrument from scratch after

realizing possible cultural incompatibilities of an instrument which was originally used (Wong, 1993).

9. Dependence on software packages

There are quite some misconceptions regarding statistical analysis. A common one is the assumption that distributions are normal / Gaussian. It is similarly not right to assume that when the data is large enough it must be Gaussian or near-to-Gaussian. The theorem that 'repeated independent Bernoulli trials tends to Gaussian' is, we would argue, mis-interpreted. One possible reason why data sets might be easily claimed to be Gaussian in nature is that it would then facilitate statistical inference. Otherwise, useful and popular statistical methods such as regression analysis and ANOVA cannot be applied to analyse the collected data. Another possible reason is our suspicion (and fear) that increasing number of researchers may not know how to test if the data collected is Gaussian, and/or to try curve fitting as an alternative to linear regression. Perhaps, more and more of us are putting our faith onto computer software and not questioning if particular analyses are meaningful or plausible. Many of us might also not know other means of analysis if our data do not fit the assumptions of these software (e.g. being Gaussian). Probably due to this reason, not too many research studies consider skewness, kurtosis or even Rasch modelling as part of the data analysis process. If you understand competence tests (mathematics test scores, say), evaluating their reliability indices is very odd.

10. Concluding remarks

In this paper, we acknowledge that student learning is a very complex process / phenomenon. Just like the Indian fable, 'blind men and the elephant', educational research often touches on part of the 'reality' of it only. Not surprisingly, the range of research will reflect different paradigms, approaches, beliefs, etc. If the phenomenon of school education is messy, so is the range of ways of researching it. In this paper, it has been our intention to alert / remind readers of the pitfalls and bridges of designing, conducting, analyzing and interpreting educational research.

This knowledge can better position us to develop a holistic image of what an elephant looks like, so to speak. Yet, at the same time, the elephant is not a static object; it develops and evolves just as we try to map out what it looks like. So does the 'reality' of student learning. For example, students' attitude to schooling changes over time. So do the conditions under which a student learns effectively. Thus, the research communities must not only work hard together to inform us how students learn and how they can be supported to learn better, they must not stop doing so because students 'evolve, classrooms 'evolve', and schools 'evolve' (see Seah, Wong, & Sum, [in press](#), for an example). The educational environment and the society change over time too.

Given how we are living in the age of the Fourth Industrial Revolution, perhaps artificial intelligence might be harnessed to complement our ongoing efforts to design, conduct, analyse and interpret educational research. But this will have to be a topic for another discussion.

References

- Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics anxiety: What have we learned in 60 years?. *Frontiers in Psychology, 7*, 1-16.
- Kagan, S. M. (1992). Professional growth among preservice and beginning teachers. *Review of Educational Research, 62*, 129-169.
- Leder, G. C. (1992). Mathematics and gender: Changing perspectives. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (A project of the National Council of Teachers of Mathematics). NY: Macmillan.

- Leu, Y.-C., Chan, Y.C., & Wong, N. Y. (2015). The relationship between religious beliefs and teaching among mathematics teachers in the Chinese mainland, Taiwan and Hong Kong. In L. Fan, N. Y. Wong, J. Cai, & S. Li (Eds.), *How Chinese teach mathematics: Perspectives from insiders* (pp. 653-699). Singapore: World Scientific.
- Ma, X. (1999). A meta-analysis of the relationship between anxiety towards mathematics and achievement in mathematics. *Journal for Research in Mathematics Education*, 30(5), 520-540.
- Marsh, H. W., & Parker, J. W. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology*, 47(1), 213-231.
- Niss, M. (2018). The very multi-faceted nature of mathematics education research. In E. Bergqvist, M. Österholm, C. Granberg, & L. Sumpter (Eds.), *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 35-50). Umeå, Sweden: PME.
- Seah, W. T., Wong, K. L., & Sum, E. S. W. (in press). Mathematics learning convictions valued by ethnic Chinese and ethnic minority students in Hong Kong: A preliminary study. In *Mathematics education research: Impacting practice (Proceedings of the 42nd annual conference of the Mathematics Education Research Group of Australasia)*. Perth, Australia: MERGA.
- Tan, C. S., & Li, Q. (2015). Affect and creativity: An old topic and new direction. In A. G. Tan, & C. Perleth (Eds.), *Creativity, culture, and development* (pp. 63-81). Singapore: Springer.
- Taylor, P. C., Fraser, B. J., & Fisher, D. L. (1997). Monitoring constructivist classroom learning environments. *International Journal of Educational Research*, 27(4), 293-302. doi: 10.1016/S0883-0355(97)90011-2
- Vale, C. & Bartholomew, H. (2008). Gender and mathematics: Theoretical frameworks and findings. In H. Forgasz, A. Barkatsas, A. Bishop, B. Clarke, S. Keast, W. T. Seah, & P. Sullivan (Eds.), *Research in Mathematics Education in Australasia 2004-2007* (pp. 271-290). Rotterdam, The Netherlands: Sense Publishers.
- Wilder, R. L. (1952). *Introduction to the foundations of mathematics*. NY: John Wiley & Sons.
- Wong, N. Y. (1993). The psychosocial environment in the Hong Kong mathematics classroom, *The Journal of Mathematical Behavior*, 12, 303-309.
- Wong, N. Y., Marton, F., Wong, K. M., & Lam, C. C. (2002). The lived space of mathematics learning. *Journal of Mathematical Behavior*, 21, 25-47.
- Zhang, Q. P., & Wong, N. Y. (2015). Beliefs, knowledge and teaching: A series of studies among Chinese mathematics teachers. In L. Fan, N. Y. Wong, J. Cai, & S. Li (Eds.), *How Chinese teach mathematics: Perspectives from insiders* (pp. 457-492). Singapore: World Scientific.