Journal of Taibah University Medical Sciences (2017) 12(2), 97–101

Taibah University

Journal of Taibah University Medical Sciences

www.sciencedirect.com

Educational Article

Revisiting the quantitative–qualitative-mixed methods labels: Research questions, developments, and the need for replication

Jimmie Leppink, Ph.D.

Department of Educational Development and Research, School of Health Professions Education, Maastricht University, Maastricht, The Netherlands

Received 25 October 2016; revised 24 November 2016; accepted 29 November 2016; Available online 5 January 2017

الملخص

إن الهدف من العلم هو وضع القوانين والمبادئ التي يمكن أن تساعدنا على تفسير الظواهر في عالمنا والكون بطريقة منهجية وفي كثير من الحالات، كيف يكون لنا القدرة على التنبؤ أو التأثير على هذه الظواهر. في هذا الشأن، يمكن لطرق البحث الكمية والنوعية إن توفر لنا أدوات المساعدة. لكن، هذه الطرق أُعطيت عدد من المسميات غير البناءة التي على الرغم من النظر إليها كركائز مفيدة يمكن أن تؤدي إلى سوء اختيارات للطرق المستخدمة في دراسة معينة. هذا المقال يبحث بعضا من هذه المسميات والجدال الذي كثيرا ما يواجهنا، ويمكن أن يسهم في الانقسام النوعي الكمي المستمر كما شهدناه في مجال التعليم الطبي ولكن غير البناء لممارسة العلم. ولوضع القوانين والمبادئ، نحن بحاجة إلى دراسات علمية مصممة بشكل جيد وتكرار هذه الدراسات. أيا كانت الطرق التي نستخدمها للتكرار المناسب، نحن بحاجة إلى توثيق جميع الخيارات والقرارات التي اتخذت طوال فترة الدراسة.

الكلمات المفتاحية: بحث كمي؛ بحث نوعي؛ طرق البحث المختلطة؛ التكرار

Abstract

The goal of science is to establish laws and principles that can help us explain phenomena in our world and universe in a systematic manner and, in many cases, how we may be able to predict and/or influence these phenomena. In this endeavour, qualitative and quantitative research methods can provide us with useful tools. However, these methods have been assigned several unconstructive labels that, although perceived as useful anchors, can result in ill-founded choices of methods used in a study. This article discusses several of these frequently encountered labels and argues that they may contribute to a continued quantitative–qualitative divide, as we have witnessed in the field of medical education, but are not constructive for the practice of science. To establish laws and principles, we need well-designed scientific studies and replications of these studies. Regardless of which methods we use, to enable replication, we need to document all choices and decisions made throughout a study.

Keywords: Mixed-methods research; Qualitative research; Quantitative research; Replication

Introduction

The goal of science is to establish laws and principles that can help us explain phenomena in our world and universe in a systematic manner and, in many cases, how we may be able to predict and/or influence these phenomena. In this endeavour, qualitative and quantitative research methods can provide us with useful tools.[1] However, these methods have been assigned several unconstructive labels that, although perceived as useful anchors, can result in ill-founded choices of methods used in a study. This article discusses several of these frequently encountered labels and argues that they may contribute to a continued quantitative–qualitative divide, as we have seen in the field of medical education, but are not constructive for the practice of science. To establish laws and

Corresponding address: Department of Educational Development and Research, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands.
E-mail: jimmie.leppink@maastrichtuniversity.nl

Peer review under responsibility of Taibah University.

ELSEVIER Production and hosting by Elsevier

 http://dx.doi.org/10.1016/j.jtumed.2016.11.008

principles, we need well-designed scientific studies and replications of these studies. Whichever methods we use, to enable replication we need to document all choices and decisions made throughout a study.[2]

Common qualitative/quantitative labels in the medical education literature

Certain labels provide useful working definitions of qualitative and quantitative methods. For instance, quantitative research typically concerns numerical data, while qualitative research usually deals with words and only minimally with numbers.[1] However, at least four types of frequently encountered labels in the literature are somewhat misleading. First, a common qualitative–quantitative labelling is that of quantitative research assuming a single truth and qualitative research assuming multiple truths. Second, quite some researchers perceive qualitative research as exploratory and quantitative research as confirmatory. Third, there is a widespread belief that in quantitative research most of the due diligence and thinking occurs prior to data collection whereas in qualitative research most of that occurs after data collection. Fourth, growing numbers of researchers appear to hold the view that mixed-methods research is always better than qualitative or quantitative alone. The following paragraphs discuss each of these types and related popular beliefs that partly result from these types.

Single truth vs. multiple truths

The first type of qualitative–quantitative labelling that is not very useful is that qualitative research assumes multiple truths while quantitative research assumes a single truth. This simplified distinction comes forth from a more useful but also more fundamental and philosophical distinction between constructivism, in which multiple realities exist, and positivism, which assumes a single reality.[3] Whether we use a qualitative or a quantitative approach, different realities or truths can be represented in different working models. For example, researchers' beliefs and experiences may influence the questions that are asked to interviewees or survey respondents, as well as how the responses are understood (cf. constructivism). In the case of qualitative analysis, the coding of narrative information into themes cannot be separated from the beliefs and experiences of the researchers who are engaged in qualitative analysis. Likewise, in the case of quantitative analysis, model choices such as whether to treat a latent (i.e., not directly observable but only indirectly measurable) variable such as knowledge as continuous or discontinuous or whether to treat a population parameter of interest as fixed (i.e., static) or as something that can vary (i.e., dynamic) are to a certain degree also a matter of beliefs and experiences.

Although researchers across medical education largely associate quantitative research with Frequentist null hypothesis significance testing,[4] where we test null hypotheses on data from random samples drawn from a population in which our parameter of interest (e.g., an average or a correlation) is fixed, quantitative research is considerably broader. Adopting a Bayesian approach,[5] for instance, different truths can be incorporated into the analysis in the form of different prior distributions. A *prior* distribution is a probability distribution regarding a parameter of interest before studying the data from a given study. For example, under the null hypothesis of 'no difference,' a prior distribution for the difference in average exam score between two groups under comparison may be a normal or otherwise symmetric unimodal distribution with more density around '0' and less density for scores further away from '0'. As data from a new study becomes available, the prior distribution is updated to the *posterior* distribution, which in turn serves as a prior distribution before data from a next study is available. Even if one is inclined to assume a single reality or truth, in the Bayesian approach that truth can vary. While in the Frequentist approach, the population parameter of interest is assumed to be fixed, and samples are assumed to be drawn randomly, in the Bayesian approach, the population parameter itself is a random variable, and samples, once observed, are fixed.

Moreover, if qualitative research were all about multiple realities or truths, the forensic science practices that deal with words and language would not have a profound impact in the criminal court case arena. Ultimately, potential multiple realities would render it utterly arbitrary to sentence a suspect who may have committed a crime in one reality but not in other realities. All qualitative analysis that is carried out in the context of forensic science to support decision-making in criminal cases is done from the starting assumption that there is one reality in which a suspect may or may not have committed a crime. Although based on the evidence available, the prosecutor may hold a plea for the *guilt* of the suspect while the defence may provide an alternative reading of the evidence under which the suspect is *not guilty*, in the end all criminal law practice assumes a single reality in which the suspect is either guilty or not. Likewise, when adopting a qualitative research approach in the study of the social construction of cancer-related fatigue experience,[6] there is no reason to assume that there must be multiple realities; it is probable that the full variety of cancer-related fatigue experiences exists in a single reality.

Unfortunately, the heuristic of equating a qualitative–quantitative distinction with that of a multiple–single truths distinction is closely linked with the popular belief that replication research has relevance for quantitative research only. In fact, the usefulness of replication research has not rarely been narrowed down even further to repeating randomised controlled experiments. Replication is fundamental to science, and if science were all about randomised controlled experiments, it would be difficult for astronomers to claim that they are practising science.

While it may be more difficult to document all choices and decisions made throughout a study in qualitative research than in quantitative research, the Reproducibility Project carried out in the field of psychology[7] clearly underlines that science, without replication, has little credibility. Whether we are interested in evidence-based practices for the design of instruction or assessment in medical curricula or in cancer-related fatigue experiences, the stakes are high. Thus, we need to replicate studies to ensure that we inform future research and practice appropriately. In qualitative studies, researchers often decide to stop collecting additional data

once they believe saturation has been reached, that is: no new information to process from additional data (e.g., interviews). When all choices and decisions made throughout a study are well documented, replication of the study with a very similar group of participants provides a straightforward approach to examine how realistic the saturation assumption was in the study that was replicated.[2] If saturation was achieved, one might expect that a replication of the study with a very similar group of participants would result in very similar findings. If the replication study leads to substantially different findings, this would provide evidence against the saturation assumption made by the researchers in the initial study. Whether we are employing qualitative methods to learn more about cancer-related fatigue experiences, experiences of anxiety in a population of patients[2] or dealing with another question in the medical or medical education domain, the stakes are high; hence, replication cannot be omitted unless researchers explicitly refrain from intentions to generalise the findings of a given study (i.e., this set of subjects, this point in time), to a broader context (i.e., subjects not studied or different points in time when phenomena of interest may have changed). However, the latter exception does not pertain to qualitative research exclusively. Not all quantitative studies have the aim or data to generalise the findings beyond the study.[4]

Exploratory vs. confirmatory

A second type of qualitative–quantitative labelling that cannot be anchored in reality is that qualitative research is exploratory, while quantitative research is confirmatory. To start, there are quantitative methods that are inherently exploratory, such as principal components analysis, exploratory factor analysis, *k* means cluster analysis, and data-driven (e.g., backward, forward, and stepwise) selection methods for the inclusion of predictor variables in regression analysis. Moreover, with the advent of eye tracking,[8] analytics,[9] and Big Data,[10] which can provide numerous measures that have a wide variety of implications for research and practice, the sky is the limit. Simultaneously, clinical and forensic research can provide examples of qualitative studies that have a confirmatory character. In forensic research, for example, one frequently starts with a limited number of competing hypotheses to then evaluate the available evidence (e.g., DNA match and eyewitness testimonies) in light of these competing hypotheses.[2,11] Likewise, studies that employ an interviewing method to learn more about cancer-related fatigue experiences may formulate their hypotheses and subsequent interview questions based on the existing literature.[6]

Equating qualitative–quantitative labelling with exploratory-confirmatory labelling facilitates the belief that qualitative research concerns hypothesis generation while quantitative research is about hypothesis testing.[1] Eye tracking, analytics, text mining,[12] and Big Data can provide large quantities of qualitative and quantitative data that can be used to generate and test hypotheses. With developments in technology (e.g., global positioning system or GPS, and social media), we have access to entire populations of interest more easily than ever before, and qualitative–quantitative and exploratory-confirmatory distinctions may fade as quickly as that of generalisation, which is commonly associated with quantitative research, *vs.* contextualisation, which is typically linked to qualitative research. Even in the still vast majority of quantitative studies that deal with a sample, generalisation is not always the goal and does not always make sense either.[4,13–15]

Due diligence before vs. after data collection

A third type of qualitative–quantitative labelling that may not be constructive for the practice of research is the idea that in quantitative research, most of the thinking occurs prior to data collection, while in qualitative research, most of this happens after data collection. While it is true that when conducting a quantitative study with particular research questions in mind, careful designing and planning may help researchers to avoid difficulty when analysing data and reporting on the study (e.g., uninterpretable findings due to heavy confounding),[16] it is also true that quantitative studies in medical education are becoming more complex (e.g., multiple measurements from the same participants and/or participants nested within centres[17]). With the advent of eye tracking, analytics, and Big Data, we may discover unanticipated findings that our research questions and study design did not consider. Moreover, the data may indicate that some of the assumptions underlying the chosen data analytic methods are violated, hence researchers need to carefully consider how to proceed with data analysis.

Concurrently, researchers who adopt a qualitative approach and do their thinking, planning, and designing work before collecting data[18–22] may avoid unnecessary trouble after data collection. Although this is not to say that keeping track of and documenting all choices and decisions made throughout a study is as difficult in quantitative as in qualitative studies, the belief that in quantitative research most of the thinking occurs prior to data collection appears to downgrade quantitative data analysis to merely clicking buttons and copy-pasting numbers. Simultaneously, the belief that in qualitative research most thinking happens after data collection has been completed may encourage researchers to 'just get started' and find themselves thinking about the consequences of not considering some important factors once it is too late.

The perceived difference in thinking, planning, and study design in qualitative *vs.* quantitative research is linked to the popular belief that issues that are related to third variables – confounding, common response, moderation, and mediation[23] – and independence of observations[17] are only a concern for quantitative research. Of course, the fact that in one case we ask a numerical response from a participant and in another case a response in words or language does not make a difference to any of the aforementioned issues having an influence. Suppose, we have seven eyewitnesses of a robbery. All seven are interviewed by the local police. The eyewitnesses differ in age, gender, profession, and other 'third' variables that may influence their testimonies. Interviewing all eyewitnesses individually before they communicate with each other may yield seven more or less independent testimonies. However, that assumption becomes problematic if the eyewitnesses either did communicate with each other before the interview or they are interviewed as a group; the effective number of

testimonies then most likely lies between one and seven and perhaps closer to one. A similar logic applies when patients, medical education residents or specialists from the same clinical centres are interviewed for an empirical study. Not only may they be more similar in experiences and response than random patients or random staff members from different centers[17] it really makes a difference whether we interview them one by one or in a group.

Mixed-methods vs. either qualitative or quantitative

A fourth type of qualitative−quantitative labelling that is becoming increasingly more popular is that mixed-methods research combines the best of qualitative and quantitative research and thus is always better than using qualitative or quantitative methods only. Amongst others, correlations in mixed-methods studies with fewer than ten participants are then justified by utilising qualitative methods as well since the latter can address 'how' and 'why' questions which often are assumed to be not addressable with quantitative methods. Actually, there is a wide range of quantitative methods that can address 'how' and 'why' questions (e.g., regression and path analysis, time series analysis, growth curve modelling, and social network analysis). However, whether any of the quantitative methods available apply depends on what research questions are addressed in a study. Some 'how' and 'why' questions lend themselves more to a qualitative approach, whereas other questions may require a quantitative approach.

While there are excellent examples in the literature of how mixed methods research has the potential to unite the best of qualitative and quantitative methods,[24] the belief that using quantitative methods on small samples is justified when qualitative methods are used as well is a misconception. Estimates of parameters of interest can be highly unstable, and both Type I and Type II error probabilities tend to be substantially elevated in the case of small samples[14,15] and reporting effect size estimates does not circumvent that problem.[25–27] For example, quantitative studies with twenty residents or fewer − which are quite common in medical education − may have such a limited statistical power that as many as 80% of the real effects are not detected in statistical significance tests (i.e., Type II errors), and the number of artefacts (i.e., Type I errors) in a group of statistically significant findings may well be more than the 5% that one would expect a priori (i.e., the conventional statistical significance level).[15] Moreover, effect size estimates may vary substantially from sample to sample,[27] for instance, from a medium size difference in favour of an experimental treatment condition in one study to a small to medium size difference in favour of the control condition in another study.

The choice of methods should be driven by the questions we wish to address as researchers, and the rules should be followed for whichever method is chosen. For quantitative methods, this usually requires a sufficient sample size.[14–16] Performing a mixed-methods study does not imply that we must collect both quantitative and qualitative data from all our participants. For instance, we may apply maximal variation sampling on quantitative scores on a response variable of interest to select which participants (i.e., the ones with scores towards the lower and upper end of the score distribution, respectively) we are going to approach for an interview.[28]

Finally, one context in which mixed-methods research is perhaps inherently useful is in that of the development of psychometric instruments. Through the combination of literature review, interviews and/or focus groups, expert panels, pretesting, and large-scale data collection and analysis, we are most likely to develop an instrument that meets the intended purpose.[29–31]

To conclude: towards more replication

When we are reminded that the goal of science is to establish laws and principles that can help to explain phenomena in our world and universe in a systematic manner and how we may predict and/or influence these phenomena, we realise that qualitative and quantitative methods can provide us with useful tools. To establish laws and principles, we need well-designed scientific studies and replications of these studies. Which methods we use should be determined by the questions we intend to address as researchers in a study, and − whichever methods we use − to enable replication we need to document all choices and decisions made throughout a study. Although they may partially stem from different philosophies, qualitative and quantitative methods have much common. Labels used in attempts to distinguish between them have not always been productive, and with recent developments related to the advent of technology, qualitative−quantitative distinctions may fade. Therefore, perhaps we should no longer think in terms of qualitative−quantitative divides but rather in terms of more-less replicable distinctions, and do all that is possible to document all choices and decisions made throughout a study to enable others to replicate our work. This will allow us to work together towards stronger conclusions and implications for future research and practice.

Source(s) of support in the form of grants

Netherlands Initiative for Education Research (NRO-PROO) Scaffolding self-regulation: effects on the acquisition of domain-specific skills and self-regulated learning skills (grant number: 411-12-015).

Authors' contribution

The author testifies that he qualifies for authorship and has checked the article for plagiarism. He conceived, designed, and carried out the literature study and wrote the full manuscript as well as the revised version of the manuscript addressing the reviewers' issues raised with the initial version of the manuscript. The author critically reviewed and approved both the initial manuscript (sent out for review) and the revised version of the manuscript (in which the reviewers' issues raised with the initial version of the manuscript have been addressed). The author is responsible for the content and similarity index of the manuscript.

Conflict of interest

The author has no conflict of interest to declare.

References

1. Cleland J. Exploring versus measuring: considering the fundamental differences between qualitative and quantitative research. In: Cleland J, Durning SJ, editors. *Researching medical education*. Chichester: John Wiley & Sons; 2015. pp. 3–14.
2. Leppink J, Pérez-Fuster P. What is science without replication? **Perspect Med Educ 2016**; 5: 320–322. http://dx.doi.org/10.1007/s40037-016-0307-z.
3. Mann K, MacLeod A. Constructivism: learning theories and approaches to research. In: Cleland J, Durning SJ, editors. *Researching medical education*. Chichester: John Wiley & Sons; 2015. pp. 51–66.
4. Leppink J, Winston K, O'Sullivan P. Statistical points and pitfalls – series – introduction. **Perspect Med Educ 2016**; 5: 42–44. http://dx.doi.org/10.1007/s40037-015-0247-z.
5. Kruschke JK, Liddell TM. The Bayesian new statistics: hypothesis testing, estimation, meta-analysis, and planning from a Bayesian perspective. http://dx.doi.org/10.2139/ssrn.2606016.
6. Graffigna G, Vegni E, Barello S, Olson K, Bosio CA. Studying the social construction of cancer-related fatigue experience: the heuristic value of ethnoscience. **Patient Educ Couns 2011**; 82: 402–409. http://dx.doi.org/10.1016/j.pec.2010.12.017.
7. Nosek BA, Aarts A, Anderson JE, Anderson CJ, Attridge PR, Attwood A, Axt J, et al. Estimating the reproducibility of psychological science. **Science 2015**; 349(6251). http://dx.doi.org/10.1126/science.aac4716.
8. Kok EM, Jarodzka H. Before your very eyes: the value and limitations of eye tracking. **Med Educ 2016**. http://dx.doi.org/10.1111/medu.13066.
9. Cooper A. What is analytics? Definition and essential characteristics. CETIS Analytics Series 1(5). [Accessed October 24, 2016] Available from: http://publications.cetis.org.uk/wp-content/uploads/2012/11/What-is-Analytics-Vol1-No-5.pdf.
10. Ellaway RH, Pusic MV, Galbraith RM, Cameron T. Developing the role of big data and analytics in health professions education. **Med Teach 2014**; 36: 216–222. http://dx.doi.org/10.3109/0142159X.2014.874553.
11. Wagenaar WA, Van Koppen PJ, Crombag HFM. *Anchored narratives: the psychology of criminal evidence*. New York: Palgrave Macmillan; 1994.
12. Toon E, Timmermann C, Worboys M. Text-mining and the history of medicine: big data, big questions? **Med Hist 2016**; 60: 294–296.
13. Leppink J, Kulasegaram KM. Avoiding common data analysis pitfalls in health professions education research. **Acad Med 2016**. http://dx.doi.org/10.1097/ACM.0000000000001448.
14. Leppink J, O'Sullivan P, Winston K. On variation and uncertainty. **Perspect Med Educ 2016**; 5: 231–234. http://dx.doi.org/10.1007/s40037-016-0281-5.
15. Leppink J, Winston K, O'Sullivan P. Statistical significance does not imply a real effect. **Perspect Med Educ 2016**; 5: 122–124. http://dx.doi.org/10.1007/s40037-016-0256-6.
16. Picho K, Artino AR. 7 deadly sins in educational research. **J Grad Med Educ 2016**; 8: 483–487. http://dx.doi.org/10.4300/JGME-D-16-00332.1.
17. Leppink J. Data analysis in medical education research: a multilevel perspective. **Perspect Med Educ 2015**; 4: 14–24. http://dx.doi.org/10.1007/s40037-015-0160-5.
18. Chen HC, Teherani A. Common qualitative methodologies and research designs in health professions education. **Acad Med 2016**. http://dx.doi.org/10.1097/ACM.0000000000001392.
19. O'Brien BC, Ruddick VJ, Young JQ. Generating research questions appropriate for qualitative studies in health professions education. **Acad Med 2016**. http://dx.doi.org/10.1097/ACM.0000000000001438.
20. Paradis E, Dan L. The tools of the qualitative research trade. **Acad Med 2016**. http://dx.doi.org/10.1097/ACM.0000000000001393.
21. Stalmeijer RE, McNaughton N, Van Mook WNKA. Using focus groups in medical education research: AMEE Guide No. 91. **Med Teach 2014**; 36: 923–939. http://dx.doi.org/10.3109/0142159X.2014.917165.
22. Teherani A, Martimianakis T, Stenfors-Hayes T, Wadhwa A, Varpio L. Choosing a qualitative research approach. **J Grad Med Educ 2015**; 7: 669–670. http://dx.doi.org/10.4300/JGME-D-15-00414.1.
23. Leppink J. On causality and mechanisms in medical education research: an example of path analysis. **Perspect Med Educ 2015**; 4: 66–72. http://dx.doi.org/10.1007/s40037-015-0174-z.
24. Myers KA, Zibrowski EM, Lingard L. A mixed-methods analysis of residents' written comments regarding their clinical supervisors. **Acad Med 2011**; 86: 821–824. http://dx.doi.org/10.1097/ACM.0b013e31822a6fd3.
25. Baguley T. Standardised or simple effect size: what should be reported? **Br J Psychol 2009**; 100: 603–617. http://dx.doi.org/10.1348/000712608X377117.
26. Cumming G. Replication and p intervals: p values predict the future only vaguely, but confidence intervals do much better. **Perspect Psychol Sci 2008**; 3: 286–300. http://dx.doi.org/10.1111/j.1745-6924.2008.00079.x.
27. Leppink J, O'Sullivan P, Winston K. Effect size – large, medium, and small. **Perspect Med Educ 2016**; 5. http://dx.doi.org/10.1007/s40037-016-0308-y.
28. Lafleur A, Côté L, Leppink J. Influences of OSCE design on students' diagnostic reasoning. **Med Educ 2015**; 49: 203–214. http://dx.doi.org/10.1111/medu.12635.
29. Artino AR, La Rochelle JS, Dezee KJ, Gehlbach H. Developing questionnaires for educational research: AMEE Guide No. 87. **Med Teach 2014**; 36: 463–474.
30. Gehlbach H, Artino AR, Durning SJ. AM last page: survey development guidance for medical education researchers. **Acad Med 2010**; 85: 925.
31. Leppink J, Van Gog T, Paas F, Sweller J. Cognitive load theory: researching and planning teaching to maximise learning. In: Cleland J, Durning SJ, editors. *Researching medical education*. Chichester: Wiley Blackwell; 2015. pp. 207–218.

How to cite this article: Leppink J. Revisiting the quantitative–qualitative-mixed methods labels: Research questions, developments, and the need for replication. J Taibah Univ Med Sc 2017;12(2):97–101.