Waikato Journal of Education

ISSN 2382-0373
Website: https://wje.org.nz

Wilf Malcolm Institute of Educational Research
Te Pūtahi Rangahau Mātauranga o Wilf Malcolm
THE UNIVERSITY OF WAIKATO

Volume 27, Issue 2 2022:
Special Issue: Navigating unexpected terrain in postgraduate research: Reflections from the field

Addressing issues of missing values in the survey research of high school mathematics teachers' digital competencies

Mairaj Jafri

Editors: Laura Gurney, Hossein Hosseini, Nicolina Newcombe and Kerry Earl Rinehart

Cite this article: Jafri, M. (2022). Addressing issues of missing values in the survey research of high school mathematics teachers' digital competencies. *Waikato Journal of Education, 27*(2), 79–84. https://doi.org/10.15663/wje.v27i2.936

Link to this volume: https://doi.org/10.15663/wje.v27i2

Waikato Journal of Education
Te Hautaka Mātauranga o Waikato

Volume 27, Issue 2, 2022: Special Issue: Navigating unexpected terrain in postgraduate research: Reflections from the field

Addressing issues of missing values in the survey research of high school mathematics teachers' digital competencies

Mairaj Jafri
The University of Waikato
New Zealand

Abstract

This paper reports how I addressed the issue of extensive missing values in my PhD study, "Digital Competencies of High School Mathematics Teachers". I collected data using an online survey. Several methods exist to address the issue of missing values. I utilised multiple imputation (MI) as it provides more accurate results. The mean scores and scale reliability of survey items changed after imputation. While addressing the missing values, I observed my focus was completely shifted from the analysis of the survey to developing an approach to imputing missing values. Researchers should be ready for complex and challenging situations. Once encountered, they should use that challenging situation to instigate creative tension – a force that moves us closer to our goals – to motivate themselves and to learn new things. I used creative tension to move from the issue of missing values back towards my initial research goal (preserving sample size and a complete dataset for analysis).

Keywords

Missing values; survey data; multiple imputation; Likert scale items; creative tension; mathematics education

Introduction

PhD students often encounter challenges in the research process despite having applied appropriate research strategies (McAlpine et al., 2020). During my PhD, I faced the issue of missing values (also known as missing data) in my data collection and analysis that consumed a significant amount of time and effort to resolve. Kang (2013) defines missing values as "the data value that is not stored for a variable in the observation of interest" (p. 402). Missing values are common in research (Farhangfar et al., 2004), and education research is no exception. If ignored, missing values can reduce the sample size available "even in initially large cohorts" (Madley-Dowd et al., 2019, p. 63) and can produce "biased estimates", leading to erroneous results (Kang, 2013, p. 403). I also discovered that the advanced statistical techniques I intended to use (such as structural equation modelling) require a complete dataset.

Wilf Malcolm Institute of Educational Research, Te Kura Toi Tangata Division of Education, University of Waikato, Hamilton, New Zealand
ISSN: 2382-0373
Contact details: Mairaj Jafri mj115@students.waikato.ac.nz
(pp. 79–84)

This situation was a form of creative tension that increased my focus on the data. I found a deep-seated desire to resolve the issue, preserve the sample size and produce a complete dataset for my research. The goal of this article is to present reflections on my journey to deal with the missing values. I aim to present how researchers can quickly decide on procedures to address missing values and move on to their main analyses with a complete data set.

Background

My PhD research aims to understand the digital competencies of high school mathematics teachers in Pakistan. Digital competence generally refers to knowledge, skills, approaches and attitudes towards digital technologies (Martin & Grudziecki, 2006). It emphasises the need for teachers to develop their digital skills (Ghomi & Redecker, 2019; Redecker, 2017) so they can search, select, organise, modify, create, download, install, share and update digital content for teaching and learning. For high school mathematics teachers it may involve an additional set of digital skills (Clark-Wilson et al., 2020; Tabach & Trgalová, 2020) including digital drawing, knowledge of equation editors, digital problem solving, computational models of mathematical objects and dynamic geometry (Dockendorff, 2020). Within these considerations, I designed and developed the Digital Competencies of Teaching Mathematics with Technology Survey.

Development of the online survey

The survey contained 36 items that were theoretically based on three constructs as defined by Tabach and Trgalová (2020): personal orientation (beliefs, attitude and preferences); mathematical digital knowledge for teaching; professional (in the classroom) and personal (administrative) use of digital resources. To articulate all dimensions, I divided the survey into six "blocks" (as shown in Table 1).

Table 1. ***Structure of the Survey***

Blocks	**Caption of each block**	**Constructs**	**No. of items**	**Measurement scale**
B1	Personal/professional information	Demographics	-	Multiple scales
B2	General beliefs about mathematics teaching	Personal orientation	7	Likert scale *Strongly agree (05)* *to* *Strongly disagree (01)*
B3	Beliefs about using DR in mathematics education	Beliefs about DR	5	
B4	Knowledge of DR & students	Instrumental geneses	24	
B5	Knowledge of DR & teaching			
B6	Knowledge of DR in curriculum planning & assessments			

*DR = Digital Resources

Blocks 4, 5, and 6 were important to measure digital competencies that are central to the learning of and doing mathematics. These blocks included statements about functions that mathematics teachers may perform with various types of digital resources; for example, the personal and professional use of digital graphs, spreadsheets, equation editors and dynamic graphing software (e.g., GeoGebra). Once the survey was created, I conducted a three-month pilot study with 42 participants to collect feedback and identify appropriate survey questions in the context of mathematics teaching with digital resources in Pakistan.

Managing expectations during Covid-19

In March 2020, during New Zealand's first Covid-19 lockdown, I realised that my data collection efforts would be limited to being done online. Travel restrictions, social distancing and school closures appeared to be impeding the possibility of obtaining an appropriate sample for analysis. I decided to modify my data collection strategy. I created a Facebook page to reach a country-wide target population of mathematics teachers in Pakistan. I also distributed the survey through an anonymous link in emails, selected LinkedIn profiles, and WhatsApp groups. Across all social networking platforms and email, the survey received a total of 306 responses. I conducted initial screening using the "completion rate" criterion, i.e., responses with less than a 10% completion rate were excluded. Following the screening process, I selected 270 responses for analysis.

Preliminary data analysis

After almost eleven months of diligence that included developing an online survey, conducting and evaluating a pilot study, and online survey distribution, the preliminary data analysis showed that a few survey items (see Table 2) consisted of a substantial number of missing values. I used SPSS 27 procedures to identify the patterns, proportion and location of the missing values. I found that 142 out of 270 cases (52.5%) were incomplete. The missing values were mainly concentrated in block 5 that was designed to measure teachers' knowledge of digital resources and teaching. Items B5_5 and B5_6 had more than 45% of values missing. It shows, that almost half of the participating teachers missed the items related to dynamic graphing software (DGS). However, the information did not tell me what caused the missing values.

Table 2. ***Summary of Items with Missing Values***

Item No.	**Items**	**Missing values**	
		n	**Percent**
B5_4	I can teach using dynamic graphing software (DGS)	25	9.3
B5_5	I can solve geometry or trigonometry problems using DGS	122	45.2
B5_6	I can use DGS for calculus and algebra	129	47.8
B5_11	I know how to screencast (screen recording), share screen and resources during video conferencing	19	7
B5_12	If required, I can teach mathematics in a fully online environment (e.g. School closure during COVID-19)	21	7.8

Why was my survey data missing?

This, I believe, was the most complex question to answer during my quest to deal with the missing values. As Jakobsen et al (2017) suggests, we never know for sure about the true nature of missing values. We can only make assumptions on how to treat data with missing values and make adjustments to the data accordingly (Graham, 2009). A researcher must determine the reasons for the missing values as the methods to deal with it depend on those reasons (Jakobsen et al., 2017). Rubin (1977) first explained that data could be missing for three reasons: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). According to Rubin (1977), when the missing values do not depend on either observed (data we have) or missing data, it is MCAR. Normally, 5 percent or less of the missing values is MCAR. Second, if the probability of missing values depends only on observed data, then the missing values is MAR. When the missing values depend on both the observed and the missing data, it is MNAR.

According to Madley-Dowd et al. (2019), it is hard to establish the relationship between how the data is missing and observed or missing values. There are some statistical tests that a researcher can perform to understand the reasons for missing values. I used separate-variance t-Tests for my study. The test helps to identify items whose pattern of missing values may be influencing the other items (Zimmerman & Zumbo, 2009). I found that teachers who answered items related to digital graphs and equation-editors were most likely to have answered items on dynamic graphing software (DGS) and video conferencing. Further, I cross-tabulated demographic items (gender) versus items with missing values. The cross-tabulation showed that the female teachers were more likely to miss items related to DGS than male teachers.

These results indicate that instead of the data being missing completely at random (MCAR), there may be a relationship between missing values and observed data. In which case, the missing values could be MAR. To test this assumption, I conducted Little's MCAR (chi-square) test, in which an insignificant value shows that the missing data is MCAR (Howell, 2018). The test results were statistically significant $(Chi - Square = 1282.818, DF = 646, Sig. = .000)$, which rejected the assumption of MCAR. The results showed that there may be a relationship between the missing values and the observed data. However, it is difficult to say for certain. The missing values pattern could have a relationship with the observed data, not with the missing data. To conclude, the data may be missing at random (MAR) given teachers' demographic and professional background. It seems that female teachers with less teaching experience, academic qualification and information technology skills are likely to skip items related to dynamic graphing software.

Coping with missing values

The next stage was imputation. In statistics, imputation is a procedure for entering a value for a specific data item where the response is missing or unusable (United Nations, 2000). There are many ways a researcher can impute missing values, including listwise or pairwise case deletion, mean substitution, maximum likelihood (ML), and multiple imputation (MI). For explanations of these methods, readers may wish to refer to the studies by Farhangfar et al. (2004) and Graham (2009). The first three methods are preferred when missing values are MCAR (5% or less), but ML and MI techniques are preferred for MAR. I preferred MI over ML for my dataset because unlike other techniques, MI replaces missing values with a "set of plausible values which contain the natural variability and uncertainty of the right values" (Kang, 2013, p. 405). I also found that MI produced more accurate estimates than the other methods (Leite & Beretvas, 2010). This does not imply that the other methods are inferior, but that I found the MI results were more representative of my dataset. After imputation, the scale-reliability score (value of Cronbach Alpha) and correlation among the survey items improved. However, the mean scores of three items related to dynamic graphing software slightly decreased after imputation. The change was expected due to a large proportion of missing values in items related to DGS. The drop did not impact the interpretation of the items' results. The mean scores of the original data showed teachers somewhat agreed that they could use DGS for teaching geometry, calculus and trigonometry, which can be interpreted as being the same for the complete dataset after MI.

What does it mean for other researchers?

At the time of writing this reflection, I recalled an anonymous quote: "When you hit the road, then you know the roadblocks." While planning, some aspects of the research design may appear small. However, they can transform into an issue whose severity can only be revealed when a researcher has to deal with them. The issue of missing values, as discussed above, illustrates the same, i.e., a common gap between initial research design and implementation reality. In other words, a gap between the researcher's vision and research reality could lead either to frustration or to "creative tension". Frustration in the sense that the focus of the researcher may shift from the main analysis to understanding the theory and the methods to deal with the issue, which might require extra work and time. However, for me, it was creative tension that I used to move from the current reality (issue of missing values) back towards the reality of the initial research vision (a complete, sparkling dataset ready for analysis).

A PhD is unlikely to be a smooth journey with all aspects "arriving" and "departing" in a timely and organised fashion. I found I needed to be ready for complex and challenging situations. Once encountered, I used those challenging situations as an act of creative tension to motivate myself and to learn new things. Chang et al. (2009) argue that "if there is no gap between the reality and the vision, the motivation process will not begin because of the lack of perceived need (and motivation) to move toward the vision" (p. 530). Senge (1990), who coined the notion of creative tension, emphasises the role of creative tension in the development of an individual. Senge argues that the ideal strategy to deal with the challenges is to move gently and continually towards your (research) goal, learning to live with the feelings of stress and emotional tension.

To constructively deal with research issues, researchers need to first understand what has happened. Some issues are common and controllable in research, such as missing values. The best way to deal with them is not to have them in the first place. For example, a more proactive approach, such as increasing the target sample size to allow for missing values or making maximum efforts to reduce missing values in the design, may be fruitful (Kang, 2013). However, in the case of unavoidable and out-of-our-control events, like Covid-19, that could affect the data collection process, we must align and adapt as soon as possible to avoid disappointment. For example, I created a Facebook page for data collection purposes that was not planned initially. Timely modifications provided flexibility to work around the pandemic and achieved the goals. We need to differentiate between situations that fall within our control and those that are beyond it. Being able to recognise the difference will help us apply solutions more appropriately. In a nutshell, researchers should keep enhancing their capacity not only to produce results but also to "master" the principles underlying the way they produce results (Senge, 1990). Doing this will generate creative tension that moves them closer to their goals.

References

Chang, Y., Eklund, T., Kantola, J. I., & Vanharanta, H. (2009). International creative tension study of university students in South Korea and Finland. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *19*(6), 528–543. https://doi.org/10.1002/hfm.20182

Clark-Wilson, A., Robutti, O., & Thomas, M. (2020). Teaching with digital technology. *ZDM - International Journal on Mathematics Education, 52*(7), 1223–1242. https://doi.org/10.1007/S11858-020-01196-0

Dockendorff, M. (2020). How can digital technology enhance mathematics teaching and learning? In R. Z. Zheng (Ed.), *Examining multiple intelligences and digital technologies for enhanced learning opportunities* (pp. 216–243). IGI Global. https://doi.org/10.4018/978-1-7998-0249-5.ch011

Farhangfar, A., Kurgan, L. A., & Pedrycz, W. (2004). Experimental analysis of methods for imputation of missing values in databases. *Proceedings of the Intelligent Computing: Theory and Applications II*, *5421*. https://doi.org/10.1117/12.542509

Ghomi, M., & Redecker, C. (2019). Digital competence of educators (DigCompEdu): Development and evaluation of a self-assessment instrument for teachers' digital competence. *Proceedings of the 11th International Conference on Computer Supported Education*, *1*, 541–548. https://doi.org/10.5220/0007679005410548

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Howell, D. C. (2018). *The treatment of missing data*. University of Vermont. https://www.uvm.edu/~statdhtx/StatPages/Missing_Data/Missing_data2x.pdf

Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials: A practical guide with flowcharts. *BMC Medical Research Methodology, 17*(1), 1–10. https://doi.org/10.1186/s12874-017-0442-1

Kang, H. (2013). The prevention and handling of the missing data. *Korean Journal of Anesthesiology, 64*(5), 402–406. http://doi.org/10.4097/kjae.2013.64.5.402

Leite, W., & Beretvas, S. N. (2010). The performance of multiple imputation for Likert-type items with missing data. *Journal of Modern Applied Statistical Methods, 9*(1), 64–74. https://doi.org/10.22237/jmasm/1272686820

Madley-Dowd, P., Hughes, R., Tilling, K., & Heron, J. (2019). The proportion of missing data should not be used to guide decisions on multiple imputation. *Journal of Clinical Epidemiology, 110*, 63–73. https://doi.org/10.1016/j.jclinepi.2019.02.016

Martin, A., & Grudziecki, J. (2006). DigEuLit: Concepts and tools for digital literacy development. *Innovation in Teaching and Learning in Information and Computer Sciences, 5(*4), 249–267. https://doi.org/10.11120/ital.2006.05040249

McAlpine, L., Castello, M., & Pyhaltö, K. (2020). What influences PhD graduate trajectories during the degree: A research-based policy agenda. *Higher Education, 80*(6), 1011–1043. https://doi.org/10.1007/s10734-019-00448-7

Rubin, D.B. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics, 2*(1), 1–26. https://doi.org/10.3102/10769986002001001

Redecker, C. (2017). European framework for the digital competence of educators: DigCompEdu. In Y. Punie (Ed.), *Joint Research Centre (JRC) science for policy report*. Publications Office of the European Union. https://doi.org/10.2760/159770

Senge, P. (1990). *The fifth discipline: The art and practice of the learning organization.* Doubleday/Currency.

Tabach, M., & Trgalová, J. (2020). Teaching mathematics in the digital era: Standards and beyond. In Y. Ben-David Kolikant, D. Martinovic, & M. Milner-Bolotin (Eds.), *STEM teachers and teaching in the digital era: Professional expectations and advancement in the 21st century schools* (pp. 221–242). Springer. https://doi.org/10.1007/978-3-030-29396-3_12

United Nations. (2000). Terminology on statistical metadata. *In proceedings of Conference of European statisticians statistical standards and studies - no. 53.* https://ec.europa.eu/eurostat/ramon/coded_files/UNECE_TERMINOLOGY_STAT_METADATA_2000_EN.pdf

Zimmerman, D. W., & Zumbo, B. D. (2009). Hazards in choosing between pooled and separate-variances t tests. *Psicologica, 30(*2), 371–390. http://www.redalyc.org/articulo.oa?id=16911991012%0AHow