ISSN 2663-8169 Vol. 5, Issue 18: 23-32, DOI: 10.25236/IJNDE.2023.051805

Investigating mathematical anxiety among school students: a discussion on different quantitative empirical procedures

Xuyang Sun

University College London, London, WC1E 6BT, United Kingdom

Abstract: *Mathematics plays a crucial role in academic development and everyday life, making it important for educators to cultivate mathematical skills in students [7]. However, there is a growing concern regarding anxiety related to mathematics among learners (ibid.). As neglecting this issue might have a negative impact on childrens' mathematical development and well-being, various educators and researchers have investigated the factors that might contribute to mathematical anxiety among school students to identify effective strategies for addressing it [11].*

Keywords: *Mathematical; anxiety; methodology; comparison*

1. Introduction

The essay reviewed two peer-reviewed research papers exploring mathematics anxiety (MA) among school students but differed in their quantitative empirical procedures. The first paper (refer to Paper 1), written by Madjar et al. [11], employed growth curve analysis to investigate the correlation between MA and various factors including gender, school transition and academic results. The second paper (refer to Paper 2), written by Hill et al. [7], mainly utilized bootstrap statistics to investigate the relationship between MA and gender differences, arithmetical performance, and general anxiety (GA).

Paper 1 is a longitudinal study which analysed how the level of MA changes over time among the same group of participants, while Paper 2 is a cross-sectional study that examines the MA of all participants at one point in time. I am interested in critically comparing and evaluating the empirical procedures in both research, such as sampling, data collection and data analysis. In specific, I would like to mainly focus on the following questions:

- Are the chosen samples appropriate?
- Are the data collection instruments appropriate?
- Are the data analysis procedures valid?

The study begins by positioning myself for conducting the study and then summarizing the two selected papers to provide a brief understanding of the research. Further, the statistical techniques adopted in both papers will be critiqued regarding the sampling process, data collection and analysis. It will be followed by a concluding summary and a reflective analysis to understand the knowledge development facilitated by the current study.

I have a strong passion for Mathematics that has encouraged me to pursue my major in the same subject to develop knowledge and skills. However, during my internship experience as a mathematics teacher, I observed that some students have considerable inhibition towards Maths, which might even lead to fearful and anxious experiences. I was aware that many students experience mathematics anxiety (MA) that makes them refrain from the subject and avoid it. Nevertheless, Mathematics is an important subject that only adds academic relevance but also contributes value to personal life through practical implementation [7]. The development of STEM education has also made it essential to develop mathematical skills among learners [7]. These findings have motivated me to explore the causes of MA and actively search for effective solutions to address this issue. Therefore, I chose two papers which investigated the factors of MA among school students to help me better understand what elements might affect students' MA.

I would like to adopt a positivist stance in my research approach. Both papers I am referring to utilized statistical analysis of quantitative data to answer their research questions, aligning with

ISSN 2663-8169 Vol. 5, Issue 18: 23-32, DOI: 10.25236/IJNDE.2023.051805

positivistic approaches. According to Somekh and Lewin [15], positivism plays a fundamental role in many contemporary statistical tests as it forms the basis for quantitative methods in social research. I agree that merely using observations and subjective experiences in society may not be sufficient enough to form solid opinions [15]. I advocate the use of samples, data, and statistical analysis. Given my educational background in Mathematics, I have learnt various statistical theories such as bootstrap statistics and ANOVA to analyze experimental results. This paper, therefore, aims to mainly evaluate the various statistical techniques employed in the two quantitative studies.

2. The two papers

2.1 Paper 1

Madjar, N., Zalsman, G., Weizman, A., Lev-Ran, S., & Shoval, G. [11]. Predictors of developing mathematics anxiety among middle-school students: A 2-year prospective study. International Journal of Psychology, 53(6), 426-432

2.1.1 Aims

The research article was published in the International Journal of Psychology in 2018. The journal is published annually with volume 58 in Issue 2 in 2023. The journal undergoes a rigorous peer review process, including editor screening, anonymous content evaluation and providing merit through independent expert reviewers to create a significant impact.

The objective of the study was to conduct a longitudinal investigation into the development of MA among middle-school students. Particularly, the study focused on the influence of gender, school transition, and scholastic achievements on the trajectories of MA over time.

2.1.2 Research design

Paper 1 adopted a quantitative study to understand the predictors of MA among middle school students. The self-reported survey was conducted with 413 students from sixth grade with 53.3% females, and a mean age of 11.27 $\pm$0.38 years. Further, the students were selected from six various Israeli elementary schools in a high socioeconomic urban area, with high-socioeconomic status according to the Israeli National Bureau of Statistics. Among the participants, 226 students underwent a transition while 187 did not.

The survey was conducted four times over the course of two years to collect data from the same population. The authors have five hypotheses before the experiment:

H1. Initial levels of MA related to gender (girls' MA might be higher), school transitions (higher for transitions) and academic performance (GPA) (higher MA level for lower GPA).

H2. MA could increase over time.

H3. The increase rate could be higher for girls compared to boys.

H4. The increase rate could be higher for pupils who transitioned schools.

H5. The increase rate could be higher for lower GPA learners.

2.1.3 Data analysis

The main analytical approach employed in Paper 1 was non-linear growth curve analysis, specifically using hierarchical linear modelling (HLM) as described by Raudenbush and Bryk [13]. According to Madjar et al. [11], this statistical technique allows for modelling the developmental trajectory of each participant and is less affected by missing data compared to other longitudinal analysis methods. By applying this approach, the researchers identified both linear and non-linear trajectories over time, which will be discussed in detail in this paper later. Additionally, they examined factors that were hypothesized to predict the initial levels of MA and the trajectories of the variables, such as gender, school transitions and GPA.

2.1.4 Findings

The researchers have addressed the five hypothesises based on the statistical analysis, as summarised below:

School transition and gender were not significantly associated with the initial level of MA, whereas

ISSN 2663-8169 Vol. 5, Issue 18: 23-32, DOI: 10.25236/IJNDE.2023.051805

the initial level of MA is higher for students with low GPAs.

MA did not significantly increase over time.

Rate of increase of MA is higher among girls compared to boys.

Rate of increase of MA is higher for students who transitioned between schools.

Rate of increase of MA was not significantly associated with GPA.

2.1.5 Ethical considerations

Authors have obtained the prior consent of the parents, guardians and principals of the target schools before collating quantitative research data from the middle school students. The study does explicitly mention that the survey data was anonymised using a unique code to prevent the identification of the students by research and educational personnel.

However, the researchers were not permitted access to school records. Therefore, students' school GPA is obtained by asking the students to indicate the range of the grades they typically received (scaled from 1-4).

2.2 Paper 2

Hill, F., Mammarella, I. C., Devine, A., Caviola, S., Passolunghi, M. C., & Szűcs, D. (2016). Maths anxiety in primary and secondary school students: Gender differences, developmental changes and anxiety specificity. Learning and individual differences, 48, 45-53.

2.2.1 Aims

The article was published in Learning and Individual Difference Journal in 2016 in volume 48. The journal is focused on publishing articles on individual differences associated with learning in a learning context (Sciencedirect.com, 2023). The journal has a double anonymised review process with an impact factor of 3.897 and a citation score of 6 (Elsevier.com, 2023).

The study examined primary and secondary students' Math and reading performance and their MA and general anxiety (GA). The authors focused on the influence of gender differences and arithmetical performance on MA. They further explored the correlation between mathematic anxiety and reading performance in tandem with the correlation between general anxiety and mathematic anxiety.

2.2.2 Research design

The research has been conducted through a survey to collect quantitative data from 981 students in Italy attending primary and secondary schools. The primary sample had 639 grade 3rd to 5th children with 322 girls at the mean age of 9 years and 5 months and 317 boys at the same mean age. The secondary sample had 342 grade 6th to 8th students with 148 girls at the mean age of 12 years and 7 months and 194 boys at the mean age of 12 years and 8 months.

The tests were conducted in the school setting, with children being tested in the classroom during group sessions that lasted approximately 1 hour each. The materials were administered in a specific order, starting with the arithmetic test, followed by the reading comprehension test, RCMAS-2, and AMAS.

2.2.3 Data analysis

Paper 2 employed a combination of bootstrap statistics, parametric tests, and regression models to assess group differences, effect sizes, and predictors of arithmetic scores. According to Hill et al. [7], the primary statistical analysis technique used in this paper was robust, distribution-independent bootstrap statistics as they employed 10,000 permutations with replacement to estimate bootstrap confidence intervals. Significant differences were determined based on whether the 95% bootstrap confidence intervals overlapped or not.

In addition to bootstrap analysis, parametric tests were also conducted, specifically 2x2 ANOVAs, with factors of gender and level of schooling. Then, to assess the relative importance of predictors of arithmetic scores, the researchers used bootstrapping to derive regression models. Variance Inflation Factors (VIF) were computed to assess the potential multicollinearity of variables in correlation tables and regression models.

For all statistical analyses, Paper 2 converted children's raw scores in the arithmetic and reading

ISSN 2663-8169 Vol. 5, Issue 18: 23-32, DOI: 10.25236/IJNDE.2023.051805

tests to standardized values. The analyses were performed using Matlab 2014a and Statistica 11 software.

2.2.4 Findings

According to the statistical analysis, Paper 1 got the following results:

Girls exhibited higher MA than boys in both primary and secondary school levels

A significant negative correlation between the arithmetic performance of secondary students and MA was revealed.

MA was moderately correlated with GA.

When GA is not considered, MA was significantly correlated with arithmetic performance.

MA was not related to reading performance GA was controlled.

2.2.5 Ethical considerations

The researcher has adhered to ethical compliance for conducting the research. In addition, formal consent has been obtained from the parents of the participating students before involving them in the research process for collecting quantitative data. However, due to a lack of parental consent, the authors were unable to administer the reading comprehension task to primary school pupils in their study. Thus, only secondary school students completed the reading comprehension test.

3. Critique

Papers 1 and 2 have a similar research aim to explore the factors of Mathematical anxiety among school students. Although both papers employed quantitative research to collect data, they utilised different statistical methods. Paper 1 conducted a longitudinal investigation, whereas Paper 2 did a cross-sectional study, namely exploring a sample at a specific time [6]. Paper 1 focused on how gender, school transition, and GPA affect MA over time, while Paper 2 focused on the influence of gender, arithmetical performance, reading performance and GA.

In addition, the two papers employed different statistical analysis techniques to get results, Paper 1 employed non-linear growth curve analysis to record and evaluate the change of MA for each participant over time, but Paper 2 mainly used bootstrapping to resample the single dataset to create many simulated samples. Although both papers reach a similar conclusion regarding the gender difference in MA, suggesting that girls may experience higher levels of MA compared to boys during their schooling, there are notable differences in their respective findings, which would be discussed in the following sections.

3.1 The quality of quantitative research.

According to Heale and Twycross (2015), critiquing involves evaluating not only the study's results but also the rigour of the research process, which refers to the measures taken by researchers to enhance the quality of their studies. In quantitative research, this is typically accomplished by assessing the validity and reliability of the study's measurements (ibid.).

Validity in a quantitative study refers to the degree to which a concept or construct is accurately measured, namely, assessing whether the measurement truly captures what it intends to measure [8]. Validity in social research can be classified into different types such as internal and external validity, criterion-related validity, content validity, and construct validity (Carmines & Zeller, 1979; Davison et al., 2006). Reliability refers to the consistency of results obtained through repeated trials or measurements [8].

Internal validity, as emphasized by Carmines and Zeller (1979), pertains to the research design's ability to attribute findings solely to the independent variable(s). On the other hand, external validity examines the extent to which the research findings can be generalized beyond the study context [14].

This paper aims to discuss the quality of different quantitative empirical procedures conducted in Papers 1 and 2. In specific, I would like to compare and critique the sampling process in both papers. After that, I would like to evaluate the quality of the questionnaires in the data collection processes. Finally, I would like to discuss the validity and reliability of the different statistical analysis techniques employed in the studies. The main focus will be on the following criteria:

ISSN 2663-8169 Vol. 5, Issue 18: 23-32, DOI: 10.25236/IJNDE.2023.051805

(1)Appropriateness of chosen samples regarding the sample size and sampling methods.

Question 1: Are the chosen samples appropriate?

(2)Adequacy and appropriateness of data collection instruments: the questionnaires.

Question 2: Are the data collection instruments appropriate?

(3)Validity and reliability of the statistical analysis techniques in concluding.

Question 3: Are the data analysis procedures valid?

3.2 Are the chosen samples appropriate?

To compare and evaluate the sample size and the validity and reliability of sampling, I summarized the sampling information in Papers 1 and 2, as listed in Table 1.

Table 1: Sampling in Papers 1 and 2

Sampling unit	Paper 1	Paper 2
School	Six selected Israeli elementary schools in a high socioeconomic urban area, with high-socioeconomic status. Three schools were "transition" and three schools were "no-transition".	Primary and secondary schools in Italy.
Duration	Follow the sample for 2 years	N/A
Students	413 participants (53.3% female) with 226 students underwent a transition while 187 did not.	639 students (50.4% female) in primary school
		342 students (43.3% female) in secondary school
Year group	6th grade, mean age of 11.27	The primary sample: grade 3rd to 5th, mean age of 9.5
		The secondary sample: grade 6th to 8th, mean age of 12.7

3.2.1 Sample size

According to Andrade (2020), a small sample size may lack the statistical power required to accurately detect differences in response, potentially leading to misleading conclusions. In the case of the research topic in Paper 1 "the development of MA among middle-school students" (Madjar et al., 2018), the sample size of 226 students who underwent a transition seems relatively small. Specifically, the findings from a sample of 226 participants may not fully generalize to all middle-school students in Israel [1]. In contrast, Paper 2 addressed this limitation by employing a significantly larger sample size, involving approximately 981 students. However, the representativeness of samples might become another important factor to consider when evaluating the quality of the research [10]. This would be further discussed in the following section.

Overall, I believe that the smaller sample size in Paper 1, particularly when examining the transition subgroup, appears to threaten the reliability of the results [8]. The limited number of participants might not adequately capture the diverse characteristics of the development of MA among middle-school students [3].

3.2.2 The validity and reliability of sampling

In Paper 1, students were sampled from six different elementary schools in a metropolitan area in Israel. The authors claimed that these schools were purposefully selected as they were geographically close and had similar characteristics. Madjar et al. (2018) suggest that children educated in urban cities may have a higher likelihood of experiencing anxiety compared to those from non-urban areas. It is possible that the authors purposely sampled schools from the urban area to account for this factor. As a result, the trajectories of MA during the 2 years might be observed more clearly. However, the reliability of Paper 1's findings may be affected due to potential confounding factors related to the school's geographic location and economic background (Psycharakis, 2011). Therefore, I think people might be cautious to interpret the findings of Paper 1, as the results may not accurately represent the broader population of middle-school students in Israel.

Despite the claim of similar characteristics among the selected schools, there may still be significant

ISSN 2663-8169 Vol. 5, Issue 18: 23-32, DOI: 10.25236/IJNDE.2023.051805

variation in the characteristics of the sampled students. Authors in Paper 1 did not clarify how they chose students in the six schools, which might therefore threaten the internal validity suggested by Carmines and Zeller (1979), as the participants may have contrasting external variables.

By contrast, Paper 2 only stated that the sampled schools are located in Italy without any other specific details regarding the sampling procedure. As a result, this paper cannot ascertain whether the selected students accurately reflect the broader population of Italian pupils.

Notably, an observed sampling bias related to gender is evident in the paper. I have calculated that the girls made up 50.4% of the primary cohort and 43.3% of the secondary cohort. The difference is statistically significant under the t-test with a 5% significance level ($p=0.034 < 0.05$), which means a significant difference from a statistical perspective. The significant difference in gender proportions between the primary and secondary cohorts suggests that the sample may not be representative of the entire population (Kim & Park, 2019). This imbalance in gender representation might lead to bias and impact the internal validity of any conclusions drawn from comparing the characteristics or outcomes of each gender group [15].

Overall, the lack of detailed information about how the participants were chosen in both papers seems to raise concerns about the validity of both studies [1]. Further, although Madjar et al. (2018) clarified the purposeful sampling of schools to examine changes in MA over two years, the small sample size and absence of information regarding student selection methods might limit the applicability of the conclusions drawn from this paper. I think this lack of transparency in participant selection methods poses a concern for the reliability of Paper 1. In contrast, Paper 2 has a larger sample size than Paper 1, but it still lacks validity due to an imbalance in gender representation. The statistically significant difference in gender proportions between the primary and secondary cohorts indicates sampling bias, which can undermine the internal validity of the research [15]. To ensure the validity and reliability of future research in this area, transparent and rigorous sampling methods might be provided.

3.3 Are the data collection instruments appropriate?

Papers 1 and 2 utilized different sets of questionnaires to quantify participants' Maths anxiety. Similarly, the questionnaires in both studies involved the Likert scale, which enabled the researchers of both papers to collect and analyse data based on fixed responses quickly and efficiently *[16]*. For example, Hill et al (2016) asked the participants to indicate how anxious they would feel during certain situations in Mathematics with a Likert scale of 1 to 5, with 1 representing the lowest anxiety and 5 representing the highest anxiety.

According to Taherdoost (2019), however, Likert scale questionnaires are not suitable when collecting quantitative data from children as they often face challenges in differentiating between the response options based on quantitative terms. The author highlighted that children have a limited understanding of the Likert response format which is highly likely to create errors in the data collection process. Therefore, as both papers involve young participants, the use of the Likert scale questionnaire might not be suitable for gaining precise and appropriate quantitative data [14]. In other words, the data reported by the participants might not truly reflect their genuine thoughts because of different interpretations of the degree of the Likert scale [14].

Another similarity between the two papers is that Cronbach's alpha was used to access the reliability and consistency of their questionnaires [5]. The Cronbach's alpha ranges from 0 to 1, where higher values indicate that the items in the questionnaire are strongly correlated with each other, suggesting that they are measuring the same construct consistently [5]. In paper 1, Cronbach's alpha was estimated to be between 0.64 to 0.79, which is less than the value reported in paper 2 (0.9). These findings suggest that the questionnaire used in paper 2 appears to have higher reliability, meaning that it is more likely to produce consistent results if administered to the same population under similar conditions [8].

However, according to Connelly (2011), Cronbach's alpha has some limitations in evaluating quantitative research. For instance, this statistics measure is influenced by the number of items in the questionnaire, with larger scales typically yielding higher alpha values. As both papers did not reveal the number of items in their questionnaire, the impact of the number of items in the questionnaire on Cronbach's alpha could not be evaluated. Therefore, I am not fully convinced that Paper 2's questionnaire can be considered more reliable solely based on the higher Cronbach's alpha value.

ISSN 2663-8169 Vol. 5, Issue 18: 23-32, DOI: 10.25236/IJNDE.2023.051805

3.4 Are the data analysis procedures valid?

3.4.1 Growth curve analysis in the longitudinal study (Madjar et al., 2018)

To track and explore the trajectories of MA over time, Paper 1 conducted a growth curve analysis technique to understand the MA among target learners over a period of two years. Growth curve analysis enables change evaluation through longitudinal data over a period of time and thus contributes to identifying variations and trends that can be administered among individuals [2]. I have acknowledged several benefits of employing growth curve analysis in this longitudinal study, such as the ability to model and compare different trajectories over time. For example, Figure 1 modelled changes in MA between low GPA and high GPA students and thus effectively demonstrated that the rate of increase in MA was not significantly associated with GPA. By utilizing growth curve analysis, the authors were able to emphasize and visually depict the observed results, contributing to a more comprehensive interpretation of the findings [11].

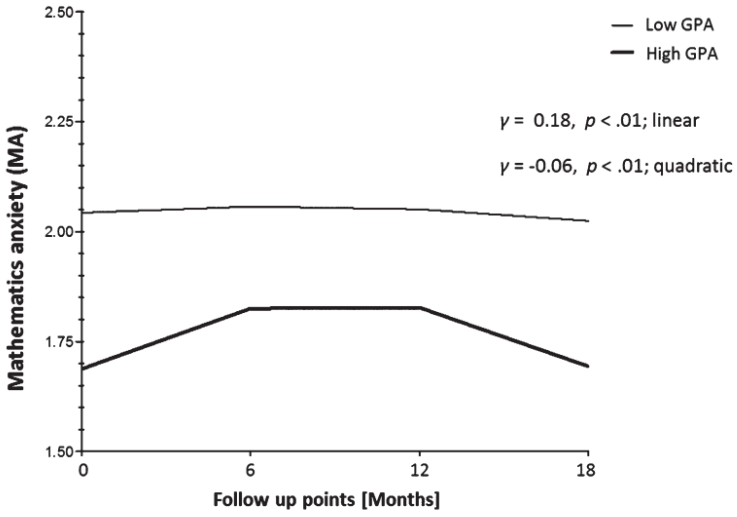

Figure 1: Growth curve of MA trajectory by GPA: low, 1 SD below the mean; high, 1 SD above the mean [11]

Although growth curve analysis is evident to provide valuable insights in longitudinal research, it is suggested to consider its limitations in analysing data and deriving findings [11]. In Paper 1, the authors categorized participants into two groups based on their GPA: high and low. However, they excluded participants whose GPA fell within one standard deviation (SD) of the mean, as shown in Figure 1. The exclusion of participants whose GPA fell within one standard deviation of the mean in the trajectory analysis might potentially threaten the internal validity of the findings[8]. In line with Carmines and Zeller (1979), internal validity refers to the extent to which the research provides evidence of a causal relationship between variables without the influence of confounding factors.

Excluding participants within a core GPA range, $(mean - 1SD,\ mean + 1SD)$, might cause a high risk of introducing bias into the analysis, as the selected sample may only involve special data among the entire population under study. This exclusion can limit the generalizability of the findings and make it challenging to draw accurate conclusions about the relationship between GPA levels and MA trajectories [3]. Therefore, I argue that the findings that the rate of increase in MA was not significantly associated with GPA might not be fully justified.

3.4.2 Bootstrap in the cross-sectional study [16]

To compare Math anxiety among different groups, Paper 2 used bootstrap statistics to construct 95% confidence intervals of Math anxiety for boys, girls, primary school and secondary school students. The 95% confidence interval means people can be 95% confident that the true average score lies within the interval [9]. For example, researchers are 95% confident that the girls' average score on the AMAS is between 21.9 to 23.1, while the boys' average score on the AMAS is between 20.0 to 21.2 (see Figure 2), which led to the finding that girls exhibited higher MA than boys. Although A 95%

ISSN 2663-8169 Vol. 5, Issue 18: 23-32, DOI: 10.25236/IJNDE.2023.051805

confidence level can ensure a low probability of extreme errors and is widely accepted in statistical analysis [4], I argue that the bootstrap strategy employed in Paper 2 might cause other errors and threaten the validity of the findings.

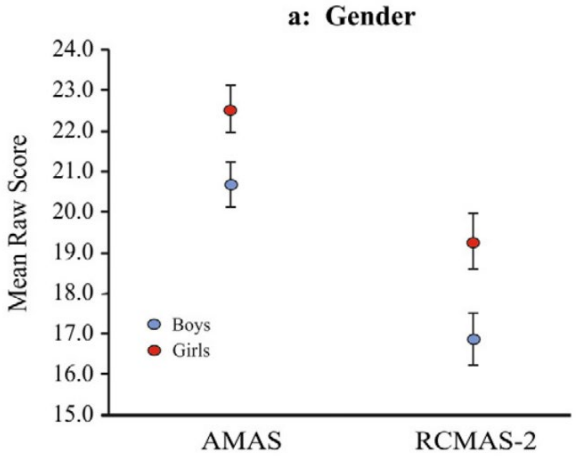

Figure 2: Mean score and 95% bootstrap confidence intervals for AMAS and RCMAS-2 scores [16]

On the one hand, Bootstrap is a non-parametric technique that does not rely on specific assumptions on underlying distributions [9]. Therefore, this technique can be applied to data that follows any type of distribution (Davison et al., 2006) [12]. According to Chihara and Hesterberg (2011), bootstrap procedures are particularly valuable when the underlying population distribution is unknown. Given that Paper 2 focuses on exploring mathematical anxiety among primary and secondary school students in Italy, which is still a relatively new research area [16], it tends to be challenging for researchers to ascertain the specific distribution that MA follows. Therefore, I agree that employing bootstrap techniques to generate additional simulated data and observe potential distributions seems to be a reasonable approach in Paper 2.

On the other hand, the bootstrap method is based on the principle that the observed sample can serve as a reasonable approximation of the population from which it was drawn [12]. It generates multiple bootstrap samples that are similar to the original sample and forms a distribution (ibid.). Therefore, a confidence interval of a specific statistic can be calculated according to the generated distribution. In other words, the accuracy of this simulation-based technique heavily relies on the original sample. If the original sample is unrepresentative, the bootstrap estimates may be biased and thus not able to provide reliable confidence intervals.

In Paper 2, the process of selection of participants was not informatively revealed. As a result, the bootstrap confidence interval reported in this paper might be questionable as the research fails to offer an adequate explanation for the sampling process and whether the participants are a good representative of students in Italy. The similar data solely generated by Bootstrap without actual students' participation might thus lack external validity to be analysed further (Davison et al., 2006).

4. Implications

Based on the findings of both papers, it is suggested to understand the reason why school students face Mathematical anxiety and conduct effective strategies to help address these anxieties. Also, as both papers revealed that girls appear to have a higher level of MA during schooling, teachers are suggested to conduct differentiated teaching methods to better support girls. Since I might become a Maths teacher in the future, in line with Hill et al's findings (2016), I would be cautious in supporting students' arithmetic performance to avoid the cause of MA.

I believe that the longitudinal study and the cross-sectional study can both be appropriate quantitative methods to achieve valuable findings based on different research aims. Additionally, I acknowledged that positivism continues to influence research design in contemporary social research, as the traditional criteria of validity and reliability remain emphasized in social science inquiry [15]. Since my dissertation will use a quantitative method, similar to paper 2, I will try to meet the criteria discussed in this paper and follow the appropriate cross-sectional procedure.

ISSN 2663-8169 Vol. 5, Issue 18: 23-32, DOI: 10.25236/IJNDE.2023.051805

5. Conclusion

According to my reviewed papers, MA tends to be a growing concern for educators as it can affect students' academic performance. Papers 1 and 2 conducted quantitative research to explore the factors of MA among school students. This paper critically discussed the two different quantitative empirical procedures in Paper 1 and Paper 2, particularly the distinct statistical analysis techniques, such as the questionnaire survey, growth curve analysis [11] and bootstrap statistics [16]. In line with my analytical findings, I would like to answer the following questions:

5.1 Are the chosen samples appropriate?

The absence of specific details regarding participant selection in both papers appears to raise concerns about the reliability of both studies. Additionally, although Paper 2 had a relatively larger sample size compared with Paper 1, the imbalance in gender representation still indicates potential sampling bias, which I believe could undermine the internal validity of the findings.

5.2 Are the data collection instruments appropriate?

I believe that both papers may have limitations due to the use of Likert scale questionnaires for young age students. Such questionnaires may not fully capture the genuine responses of young children, and students may face challenges in quantifying their answers accurately [10]. Therefore, caution needs to be exercised when interpreting the results.

5.3 Are the data analysis procedures valid?

Although both papers seem to choose a reasonable technique, there is evidence that they still have limitations. Paper 1 employed growth curve analysis to examine changing trends in MA over two years, providing clear diagrams to visualize predictors. However, the researchers excluded data within a core GPA range when modelling the curves, which may introduce bias and threaten the internal validity of the findings. Similarly, the distribution modelled by Bootstrap statistics in Paper 2 might be questionable as the study lacks information on whether the original data are a good representation of students in Italy. Thus, the use of Bootstrap to generate similar data without actual student participation raises doubts about the external validity of the findings [6].

Future research might be needed to employ a more transparent sampling process. Additionally, replication studies are suggested to address concerns regarding the internal and external validity of the results.

6. Reflections on the review process

By reviewing the two research, I have developed a better understanding of quantitative methods and positivism theory. By critiquing the sampling process, I found various difficulties in selecting samples, such as balancing external validity and internal validity. In addition, I have acknowledged that the growth curve could be supported by multiple linear regression based on the research objective and purpose, which might contribute to determining the linear impact of the independent variable on the dependent variable. Further, I also learned that although Likert scale questionnaires appear to be appropriate for the quantitative research study, it seems to be important to note that they may not be equally relevant when conducting research with children. I believe that I might consider providing some open-ended questionnaires to the participants, particularly young kids so that they might express their opinions and perspective better.

References

[1] Andrade, C. (2020). Sample size and its importance in research. Indian Journal of psychological medicine, 42(1), 102-103.
[2] Bauer, G. R., Churchill, S. M., Mahendran, M., Walwyn, C., Lizotte, D., & Villa-Rueda, A. A. (2021). Intersectionality in quantitative research: A systematic review of its emergence and applications of theory and methods. SSM-population health, 14, 100798.
[3] Carmines, E., & Zeller, R. (1979). Reliability and Validity Assessment. SAGE Publications, Inc.

ISSN 2663-8169 Vol. 5, Issue 18: 23-32, DOI: 10.25236/IJNDE.2023.051805

[4] Chihara, L. M., & Hesterberg, T. C. (2011). Mathematic statistics with resampling and R. Hoboken. NJ: Wiley.
[5] Connelly, L. M. (2011). Cronbach's Alpha. MedSurg Nursing, 20(1), 45-50.
[6] Davison, S. N., Jhangri, G. S., & Johnson, J. A. (2006). Cross-sectional validity of a modified Edmonton symptom assessment system in dialysis patients: A simple assessment of symptom burden. Kidney International, 69(9), 1621–1625.
[7] Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics anxiety: What have we learned in 60 years? Frontiers in Psychology, 7, 508. https://doi.org/10.3389/fpsyg. 2016.00508
[8] Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. Evidence-Based Nursing, 18(3), 66–67. https://doi.org/10.1136/eb-2015-102129
[9] Hill, F., Mammarella, I. C., Devine, A., Caviola, S., Passolunghi, M. C., & Szűcs, D. (2016). Maths anxiety in primary and secondary school students: Gender differences, developmental changes and anxiety specificity. Learning and Individual Differences, 48, 45–53.
[10] Kim, T. K., & Park, J. H. (2019). More about the basic assumptions of t-test: Normality and sample size. Korean Journal of Anesthesiology, 72(4), 331–335.
[11] Madjar, N., Zalsman, G., Weizman, A., Lev-Ran, S., & Shoval, G. (2018). Predictors of developing mathematics anxiety among middle-school students: A 2-year prospective study. International Journal of Psychology, 53(6), 426–432.
[12] Psycharakis, S. G. (2011). A Longitudinal Analysis on the Validity and Reliability of Ratings of Perceived Exertion for Elite Swimmers. The Journal of Strength & Conditioning Research, 25(2), 420. https://doi.org/10.1519/JSC.0b013e3181bff58c
[13] Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (Vol. 2). Thousand Oaks, CA: Sage.
[14] Taherdoost, H. (2019). What is the best response scale for survey and questionnaire design; review of different lengths of rating scale/attitude scale/Likert scale. Hamed Taherdoost, 1-10. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3588604
[15] Somekh, B., & Lewin, C. (2005). Research methods in the social sciences. SAGE Publications.
[16] Williams, J., Roth, W. M., Swanson, D., Doig, B., Groves, S., Omuvwie, M. & Mousoulides, N. (2016). Interdisciplinary mathematics education. Springer Nature. https://library.oapen. org/bitstream/ handle/ 20.500.12657/27783/1/1002222.pdf