JUTLP
Journal of University Teaching & Learning Practice

Editorial: Improving quantitative research in higher education learning and teaching

Professor Louise Taylor[a], Dr Hai Min Dai[b], Dr Hilary Ng[c], Dr David Playfoot[d], and Dr Sarah Rose[e]

[a] Oxford Brookes University, United Kingdom; [b] Shanghai Jiao Tong University, China; [c] Hong Kong Metropolitan University, Hong Kong SAR, China; [d] Swansea University, United Kingdom; [e] University of Staffordshire, United Kingdom.

Abstract

Empirical research in higher education learning and teaching is vital if we are to enhance our understanding of how to improve students' experiences and outcomes. Historically, however, this research has received less attention, recognition, and investment compared to discipline-specific research. Consequently, the field has suffered from a lack of methodological quality. In this Editorial, we consider common methodological problems in this research in view of international standards for quality research. Focusing on quantitative methods, we offer guidance to educators for developing rigorous higher education research, using a psychological science lens. We structure our discussion around three Principles for conducting theoretically-driven and systematic research, using tools that can increase the reliability, validity, and subsequent impact of research. Principle 1 = Theorise well, Principle 2 = Prioritise method robustness, and Principle 3 = Analyse, do not merely describe. Ultimately, our aim is to support educators to conduct robust research for improving higher education learning and teaching, and facilitate the dissemination of impactful findings by sharing them with others at scale in educational journals.

Editor

Editor-in-Chief: Dr Joseph Crawford
Senior Editor: Dr Alison Purvis

Publication

Submission: 14 October 2024
Revised: 7 March 2025
Accepted: 6 May 2025
Published: 26 May 2025

Practitioner Notes

1. There is an urgent need for more high-quality quantitative research in higher education learning and teaching that provides a robust evidence base for improving students' experiences and outcomes.
2. Theorise well: Theory should be used to shape research questions, justify the selection of variables, and guide methodological decisions.
3. Prioritise method robustness: Rigorous research design, with necessary controls where appropriate, will strengthen the quality of the data.
4. Analyse, do not merely describe: Conducting appropriate inferential statistics means that data can be generalised beyond the specific sample of students in the research.

Keywords

Educational research; Quantitative research, Scholarship of learning and teaching; Research Design; Research Impact

Citation:

Taylor, L., Dai, H. M., Ng, H., Playfoot, D., & Rose, S. (2025). Editorial: Improving quantitative research in higher education learning and teaching. *Journal of University Teaching and Learning Practice, 22*(2), https://doi.org/10.53761/jc59jm21.

Introduction

Research in higher education learning and teaching is essential for understanding how students learn and for identifying effective pedagogic approaches that support student success. Central to this inquiry are questions such as: How can we engage students with material that they may not see as relevant? How can we integrate new technologies to enhance learning outcomes? What assessment strategies enable all students to demonstrate their understanding? Addressing these questions is fundamental to fostering excellence in teaching and learning (Trigwell, 2013).

The academics and practitioners engaged in this research represent the rich diversity of disciplinary backgrounds, yet they share a common commitment to improving learning and teaching. They are motivated by the ultimate goal of enhancing teaching in a way that promotes meaningful learning among their students. This research tradition is often referred to as the ‘scholarship’ of teaching and learning (SoTL), following the seminal work of Boyer (1990). SoTL includes not only the systematic study of teaching and learning, or what Boyer called *discovery*, but also the application of research findings to enhance teaching practice (Boyer, 1990).

Work done as part of the *discovery* phrase of SoTL, however, has faced criticisms for being ‘the thorn in the flesh of educational research’ (Canning & Masika, 2022), largely because of concerns about its quality when compared with discipline-specific educational research (Cotton et al. 2018). A key issue is the frequent lack of theoretical grounding in higher education learning and teaching research (Pownell et al. 2025; Macfarlane, 2011; Canning & Masika, 2022). Furthermore, it is often conducted by those with limited formal training in research methods (Rowland & Myatt, 2014). For instance, Evans et al. (2015) evaluated almost 300 articles to assess the strength of the evidence base supporting pedagogic strategies for enhancing student engagement. Their findings revealed that only 13% of the articles were of high quality, with many failing to include sufficient methodological details, and many providing inadequate analysis accompanied by over-interpretation of data. Given these limitations, it is unsurprising that research underpinning SoTL is often viewed as secondary to discipline-specific research (Boyer, 1990; Waller & Prosser, 2023) and is undervalued for career progression (Cotton et al. 2018). As a consequence, the field has not received the rigour that it deserves, limiting its overall quality and impact (Canning & Masika, 2022; Evans et al. 2021; Hartas, 2015; How, 2020; Tight, 2017). Given that learning and teaching research in higher education has proliferated over recent years (Bull et al. 2024) overcoming these challenges is vital to strengthen the field.

With these issues in mind, this Editorial aims to highlight common design pitfalls in quantitative higher education learning and teaching research. Many of these pitfalls we see in manuscripts that are submitted to the Journal of University Teaching and Learning Practice. The second aim is to provide guidance on how to avoid these problems, drawing on principles of rigorous research design using a psychological science lens. In doing so, we seek to support educators to improve research underpinning SoTL, which has the direct aim of enhancing teaching practice (Boyer, 1990). Hopefully, this Editorial will also serve as a reminder for all researchers of ‘back-to-basic’ design principles, as well as support less experienced educators/researchers to avoid common pitfalls as they take their first steps into higher education learning and teaching research.

Defining Quality

To understand how to improve research quality, it is first important to consider what it is and how it can be measured. We draw on international research assessment exercises conducted in Australia (Excellence in Research for Australia, ERA, 2018) and the United Kingdom (Research Excellence Framework, REF, 2019). These exercises judge research outputs in terms of three main issues: 1) originality - new or transformative ways of thinking about a topic; 2) significance - influence in a particular field; and 3) rigour - appropriate methodology with compelling evidence that the aim has been met. Furthermore, the research impact, or the extent to which the findings have transformed practice and people, is measured according to both significance (as defined above) and 'reach', that is, the number and extent of beneficiaries who have experienced a positive change from the research. A lack of theory or methodological rigour minimises the extent to which findings can contribute new knowledge, and this substantially reduces the potential impact for positive change. Manuscripts in this field, and those submitted to JUTLP, are not always grounded in theory, but instead are often motivated by seeking to understand an immediate problem within one's own teaching practice or discipline. In addition, the methods are often underspecified or lacking in design sophistication and rigour, which limits the scope of questions that can be answered. Howard and Brady (2015) note that there is an abundance of case-study type research or 'practice papers' in this area that can lead to small-scale changes (usually in individual classrooms), but usually fail to influence pedagogic practices more widely. If we really want to quantify the impact of educational change on students and positively change practice, we need carefully designed studies that meet internationally recognised definitions of quality (Hartas, 2015).

In summary, we argue that there is an urgent need for more theoretically positioned and methodologically rigorous quantitative research, which improves learning and teaching both within *and* beyond our disciplines. In this Editorial, we aim to support researchers to achieve this by addressing common pitfalls in this area, including those we see in manuscripts submitted to JUTLP. We also draw on principles of research from psychological science to enhance quality. First, we discuss the scientific process. Second, we share three key Principles for conducting quality research. Principle 1 = Theorise well; Principle 2 = Prioritise method robustness; and Principle 3 = Analyse, do not just describe.

The Scientific Process

Epistemological stances about the nature of knowledge—such as empiricism, rationalism, and positivism—fundamentally shape research approaches in both the natural and social sciences. Empiricism emphasises that knowledge primarily derives from sensory experience, asserting that valid knowledge about the world can be discovered through direct observation. In contrast, rationalism prioritises reason and intellectual processes, advocating for conceptual clarity and the use of established categories to interpret data (Hjørland, 2005). Positivism, especially logical positivism, combining both empiricist and rationalist views, posits that knowledge can be reduced to a set of statements, and all scientific meaningful statements must either be empirically verifiable or logically necessary. This positivist-empiricist approach underscores the importance of observation and verification in uncovering the universal laws underlying phenomena (Ejnavarzala, 2019).

While there are ongoing debates regarding 1:1 corresponding relationship between epistemological stances (epistemology) and technical practices (research methods) in social sciences (Bryman, 1984), quantitative research generally follows the positivist assumption to test objective theories by examining the relationships between variables (Creswell & Creswell, 2022). Contemporary quantitative research aligns more to post-positivism, which acknowledges the existence of an objective reality while also recognising the complexity, context and limitations of human understanding.

The research process of quantitative research is theoretically driven and typically follows a deductive approach, starting from using or developing a theoretical framework or model based on existing knowledge, identifying variables, formulating hypotheses, designing measurement tools, collecting and analysing data, to validating, refining, or refuting the theories/models under investigation. Engaging in this structured process will ensure that quantitative research in higher education is rooted in empirical evidence and tested against theoretical constructs. This research can then contribute to the development of educational theory while also offering strategies or solutions to real-world challenges.

Psychological science related to higher education learning and teaching research

Educational research is complex. It is a subject that combines individual variations in both learners and teachers with social contexts within and outside the classroom. It is also impacted by the socio-political climate that determines what education can and should be (Berliner, 2002). Education research in the domain of higher education adds another layer of complexity. Students in our classrooms are in a period of significant change in their lives. This poses challenges for their mental health and wellbeing as they transition into, through, and out of university (Cage et al. 2021). As a discipline, psychology is well-placed to inform research in this area through its use of robust methods and extensive theoretical grounding.

Psychologists use scientific methods to answer questions about human behaviour and to improve people's lives. Decades of psychological research have led to the formation of robust, relevant, and evidence-based theories that can be applied to education to understand how people process, learn and retain information. Theories seek to enhance our understanding of the cognitive, emotional, behavioural, and social processes involved in learning. These may include psychological processes relating to how memory works, how we process information, what motivates us to learn (or not), and the impact of context and others on our learning. In addition, quantitative psychologists draw heavily on the scientific method to develop testable predictions about behaviour by gathering empirical evidence through well-controlled and carefully designed experiments, using statistical techniques to analyse data, and to contribute to theory development. Therefore, psychology offers a rich theoretical and scientific tradition for those wishing to conduct high quality research in higher education learning and teaching.

In the remainder of this Editorial, we draw on our shared expertise as researchers, educators, journal reviewers and editors, and psychologists, to provide three key Principles for improving quantitative research in higher education learning and teaching. These are: Principle 1, Theorise well; Principle 2, Prioritise method robustness; and Principle 3, Analyse, do not merely describe.

Three principles for enhancing quantitative research on higher education learning and teaching

Principle 1: Theorise well

In quantitative research, theories serve as foundational frameworks that shape research questions, justify the selection of variables, and guide methodological decisions. Theories provide a structure, enabling researchers to examine relations between variables. This helps to ensure that studies contribute meaningfully to the cumulative body of knowledge in the field. Theoretical frameworks help define constructs, determine appropriate measurement tools, as well as establish hypotheses that are logically derived from prior empirical findings.

Despite the central role of theory in research design, some learning and teaching research in higher education suffers from a lack of theoretical or empirical justification for the selection of variables. In such cases, variables appear to be chosen arbitrarily rather than being developed from established theoretical constructs. A notable example of this is gender. On the one hand, it is good practice to check whether any gender differences exist in the data so that these can be controlled for in analyses. Often, however, if gender differences are found, it is tempting for authors to give these a prominent focus (particularly in the Results and Discussion sections), when no a priori rationale was provided. In other cases, examination of gender differences may be a valid focus of the research, and as such, it is essential to include a solid theoretical and/or empirical basis for doing so.

As an example, Xia at al. (2023) investigated whether gender and need satisfaction impacted self-regulated learning in AI-enabled environments. They first provided a thorough review of research on gender differences in perceptions of AI and technology acceptance in learning, as well as gender differences in need satisfaction as conceptualised through self-determination theory. Building on this discussion, they then proposed several hypotheses regarding the role of gender, which they could then test empirically, and the findings could be used to build on existing data and theory. In summary, all variables being measured should be motivated theoretically or empirically.

Another common problem in higher education learning and teaching research involves superficially evidencing outcomes by simply describing the interventions used to change classroom practices in the absence of any theoretical rationale for either the intervention or the methods of data collection. For example, an educator may design an intervention to improve student engagement using virtual reality (VR) technology and find positive results. But, if the study lacks a theoretical framework to justify the choice of VR as an engagement tool, or lacks reference to relevant theories of student engagement, it will likely fail to advance our understanding of how or why VR specifically may support this outcome, or whether there is another explanation. This lack of theoretical grounding can limit the research quality and subsequent potential for real-world impact because it lacks clear rationale. Thus, without theory, future research in higher education risks adding to the collection of isolated cases of learning and teaching, rather than contributing to a broader body of educational knowledge (Howard & Brady, 2015).

Once a relevant theoretical framework has been identified, it is important to consider the following issues, which can be overlooked. Researchers need to clearly justify the relevance of their choice of theory. For example, is it the most robust framework for explaining a specific educational phenomenon? Does it provide a unique perspective to understand the phenomenon under study? Moreover, researchers also need to ensure that the data collection tools align with the conceptualisations of the constructs in the selected theory. They should also ensure that the data collected has the potential to support or refute that theory in order to facilitate the development of theory (Ashwin, 2012). As an example, there may be several theories applicable to examining the role of motivation on student performance, and a single study will understandably probably focus on one of these theories to the exclusion of others. What is important is that the researchers provide a rationale for this choice, so that the reviewers, editors, and readers can be assured that the manuscript reflects a deliberate and critical choice rather than an artefact of an inadequate literature search that missed potentially relevant sources.

In summary, research that is grounded in a theoretical framework enhances the overall quality of the study by increasing conceptual clarity, methodological rigor, and alignment with existing academic discourse. A strong theoretical foundation helps justify the selection of research variables and facilitate the interpretation of findings within broader educational theory. It also increases the likelihood that its findings are practically informative by helping us understand how we can create meaningful change in our teaching.

Principle 2: Prioritise method robustness

Use of a theoretical framework alone does not, unfortunately, guarantee that a quantitative study will be of high quality and impact. A related integral consideration is what research design is best suited to measuring the variables under investigation. It is important to highlight that there is no one “right” design to use for any one study; designs can be more or less suited to particular research questions or research contexts. In this section, we outline common research designs (experimental and survey designs, and cross-sectional and longitudinal designs) and provide descriptions of what they can (and cannot) evidence. We also consider the importance of controlling for confounding variables through careful sampling and offer practical guidance for improving research design.

Experimental design versus survey approaches

Experimental designs and survey approaches have very different purposes and methods. The principal characteristic of experimental design is that it aims to establish cause-and-effect relationships between variables. In contrast, survey studies are focused on describing similarities or differences between groups of people and identifying patterns or correlations between variables. In higher education learning and teaching research, experimental designs are essential for trying to understand ‘what works’; they are particularly suited to asking questions about the effectiveness of new teaching methods or interventions. Unfortunately, survey or correlational studies are more common in this field than experimental studies, which limits our understanding of the impacts of many educational interventions. In survey studies, authors may draw conclusions that cannot be inferred from the data collected in the way that the study was conducted. For example, it is tempting to use data obtained from a survey to argue that a change to course design has been successful, perhaps because the change correlated with higher student attendance or grades. But they are merely reporting a *relation* between a new design of course and student outcomes; this cannot be used as evidence that the design has causally impacted student outcomes. To draw the latter conclusion, an experimental design is needed (e.g., to rule out the possibility that the old design was equally successful).

Experimental design: Experiments aim to test hypotheses by manipulating independent variables and observing their effects on dependent variables. Group membership is predicated on a specific variable, for example, year of study, or whether students received a particular teaching intervention. In comparing two or more groups, it is essential that one of those groups acts as a *control* group, or 'business as usual' against which to compare performance in the *experimental* group(s). This is particularly important when evaluating the effectiveness of pedagogic interventions. In practice, control groups are often lacking in research in this field. This may be because, practically, they can be difficult to enact, but with careful consideration this can be achieved.

By way of an example of an experiment with a control group, a study by Zhao et al. (2024) examined cheating rates by comparing performance of students who were reminded of the academic dishonesty policy on a test paper (*experimental* group) versus those who were not (*control* group). Furthermore, students were assigned *randomly* to either the experimental or control group. Random assignment helps ensure that 1) groups being compared are similar at the outset, and 2) that the difference between groups is due to the intervention rather than a characteristic of the participant group. Five of the questions were fill-the-blank questions that were carefully designed by the researchers to be almost impossible for students to complete without cheating by looking-up the answers on their online learning portal during the test (although this had been explicitly forbidden during the test). Although cheating rates were concerningly high, the cheating rate in the *experimental* group (55%) was significantly lower than in the *control* group (69%). This enabled the authors to conclude that students who received the intervention cheated significantly less often than students in the control group, and that this intervention could serve as a useful teaching practice in the future.

Having control groups, however, introduces a potential complication in relation to ethics. A full consideration of the ethical issues in learning and teaching research is beyond the scope of this Editorial (see instead Purvis & Crawford, 2024), but it is important to acknowledge that, by definition, the control group will not be exposed to a variable that could be predicted to benefit learning. This creates a potential difficulty for the educator-researcher: as a researcher, we need control groups so that we can draw appropriate conclusions, but as educators we want all of our students to be exposed to something that may benefit their learning. Therefore, we may need to consider how the control group will be provided with the benefit at a later date, so as not to leave a *lasting* disadvantage. There is also the possibility that the manipulation or intervention might end up being detrimental to the experimental group, although this becomes less likely when there is a strong theoretical basis for the intervention in the first place. In any event, it is important that researchers are mindful of the potential impact of participation on the students that they study.

Survey design: Using questionnaires to evaluate students' experiences are ubiquitous in higher education learning and teaching research (Ali et al. 2021). However, common issues with their use in this field are: 1) lack of theoretical rationale for adopting a particular questionnaire, and, relatedly, lack of alignment between a research question and use of a particular questionnaire (See Principle 1: Theorise well, and Purvis, et al. 2024); and 2) creating a new questionnaire when a preexisting one already exists, alongside failing to consider issues of reliability (consistency of measurement) and validity (accuracy of measurement).

There are many reliable and valid published questionnaires, and researchers should always look to use these first before creating a new one. Designing a new questionnaire can be a research study in its own right (e.g., McAvoy et al. 2021; Vignoli et al. 2023). When selecting a questionnaire to use, it is important that there is evidence of its reliability and validity (Drost, 2011). It is also important to report the reliability of the questionnaires used in a particular study with the sample. There are a number of different approaches that can be used to assess validity and estimate reliability.

In terms of validity, the first step is usually to reflect on whether the questionnaire contains questions which cover all aspects of the construct being measured (content validity) and whether the questions appear to measure what the researchers are intending them to measure (face validity). As both content and face validity are rather subjective, a more objective measure is construct validity. This usually involves exploratory and confirmatory factor analysis (Field, 2025, provides a comprehensive guide). These statistical techniques are used to test the extent to which the data from a questionnaire provides a good representation of our theoretical understanding of the construct, thereby testing whether the questionnaire measures what it is intended to measure. In terms of reliability, probably the most commonly reported measure is Cronbach's alpha. Ideally, a questionnaire should have a Cronbach's alpha of at least .7, because this is generally regarded as the lowest value representing an acceptable level of internal reliability (Taber, 2018).

These approaches to assessing validity and estimate reliability are based on Classical Test Theory. This approach is based on relatively straightforward mathematics (e.g., averages, proportions, correlations) making it comparatively simple to understand, and it also does not require very large samples to produce meaningful results (Raykov & Marcoulides, 2016). While this approach still dominates much learning and teaching research there is an alternative approach, Item Response Theory (Sharkness & DeAngelo, 2011). This is a much more complex approach to analysing tests, with the focus being on the item, rather than overall scale level, thereby allowing for devising, revising, and optimizing scales for specific uses (De Ayala, 2009; Jabrayilov et al. 2016). The general steps involved in an Item Response Theory analysis can be found in Toland (2014), who provides a detailed guide to Item Response Theory. The key message here, regardless of whether a Classical Test Theory or Item Response Theory approach has been taken, is that it is critical to provide detail relating to why a particular questionnaire has been chosen and demonstrate that its suitability and robustness as a data collection tool has been considered.

Cross-sectional versus longitudinal approaches

Cross-sectional and longitudinal research designs can be used with both experimental and survey methods, and they are two of the main types of designs used in quantitative research. Cross-sectional designs usually involve collecting data at a single time point, perhaps from multiple groups of participants that differ on a key variable (e.g., ethnicity or disability). They offer researchers an effective way to collect data without the need for long-term follow-up, making them particularly useful when constrained by limited funding or time. However, they can only provide a snapshot of correlations between variables, and they cannot determine changes over time or establish cause-effect relationships.

Longitudinal designs involve collecting data for the same set of variables from the same participants over at least two time-points over an extended period of time. This enables researchers to describe developmental changes or intervention effects across time. For example, in a study by Thiem et al. (2023), they investigated the effectiveness of a research-based-learning intervention by assessing students' self-rated research competence across five time-points over two years. Statistical analysis was used to examine within-person changes across these time intervals, and control for whether or not they had received the intervention in each of the five time intervals. Analysing within-person changes in this manner enabled the researchers to control for individual differences and changes that may have occurred over time, irrespective of the intervention (e.g., increases in general topic understanding, communication skills, maturity etc.). This rigorous analysis enabled the researchers to be confident in their conclusion that the research-based-learning intervention led to increases in students' self-rated research competences, and not other factors. Only with this type of design can we be confident in the direction of causality and begin to rule out alternative explanations. In essence, longitudinal designs can provide a more detailed understanding of how students learn over time, and can identify patterns of change to truly understand the impact of educational interventions.

Confounding variables and sampling issues

Irrespective of the strength of the research design, it can only be as good as the quality of the sampling and ability to control for potential confounding variables. Confounding variables are those that influence participant responses but are either not of interest to the research question or have not been considered (Hartas, 2015). In higher education learning and teaching research, self-selecting samples are common but their limitations are often not recognised. The fact that many students self-select into a study is important to consider because this can introduce a confound, or bias, if the resultant sample does not represent the population. It is not uncommon, for example, to invite a whole cohort of students to take part in a study, but not every student will respond. It is likely that those who volunteer share certain characteristics, such as, they may be particularly motivated or engaged in their learning, highly conscientious, or keen to help others. The experiences and performance of these students may not reflect the whole cohort, and this can skew the results. Therefore, it is useful for researchers to include an estimation of the proportion of participants who responded from the available pool because this can provide an estimate of whether there are likely to be confounding variables. Ensuring that sample parameters are defined and that samples are as representative as possible increases the validity and real-world applicability of the data (Kanaki & Kalogiannakis, 2023). This approach not only enhances the credibility of the research but also ensures that the data reveals whether the teaching interventions and strategies are effective, regardless of students' background or circumstances.

Careful sample selection can minimise the impact of confounding variables. Truly random sampling will generally lead to the fewest confounds because participant characteristics are not likely to vary (or cluster) systematically (Creswell & Creswell, 2022). For example, imagine that an experimental study tests an intervention to explore students' preference for oral feedback (the experimental group) versus standard written feedback (the control group). The experimental group contains the first 20 students to attend class, whereas the control group comprises the remaining 20 late arriving students. Imagine the findings are that students in the oral feedback group were extremely positive about their feedback, and students in the written feedback group were not so positive about their feedback. It would be tempting to conclude that students prefer oral versus written feedback, however, arrival time acts as a confounding variable here. It could be that those students in the oral feedback (experimental) group (who were also on time) were more organised, and possibly higher performing, than those in the control group, and these factors may be associated with feedback ratings rather than the mode of feedback itself (i.e. it could be that students in the experimental group are more positive about feedback than the control group, irrespective of the mode of feedback delivery). Thus, lack of random assignment does not allow us to draw any conclusions about the method of feedback preferred by these students. The confound of arrival time could have been avoided by randomly allocating students to the two groups. Of course, truly random sampling is not always possible in learning and teaching research because potential participants are often determined by enrolment on a particular course or within a particular discipline. When random sampling is not possible, researchers ought to acknowledge the potential limitations of their sample, including potential biases and lack of representativeness.

In summary, this section has introduced some key issues in research design to support researchers to conduct high quality quantitative research. Experimental designs are best suited to measuring the true impact of teaching interventions on student outcomes, whereas correlational designs simply describe relations between two or more variables at any one time. A key message for researchers relates to the importance of controlling for confounding factors that may obscure true relations between variables, whilst also including appropriate control groups. Clearly articulated sample parameters are also essential to help readers understand the extent to which the findings are representative of the population.

Principle 3: Analyse, do not merely describe

A well-conceived method is paramount because it lays the foundation for robust data analysis, ensuring that the results are valid, reliable, and directly address the research question(s). In the following section, we discuss the importance of going beyond descriptive analysis when interpreting data and provide key considerations for researchers.

Data analysis: Moving beyond descriptive statistics to inferential statistics

Assuming you have been able to collect good quality data, the next step is to analyse it in order to address your research question. Unfortunately, simply describing the data (i.e., using descriptive statistics) is unlikely to be sufficient to draw conclusions that will be meaningful on a broader scale. Descriptive statistics simply summarize the main features of data and provide an overview of its characteristics. Descriptives include measures of central tendency (mean, median, mode) and measures of dispersion (range, variance, standard deviation). In contrast, inferential statistics go beyond describing the data and aim to make inferences or generalisations about a larger population (e.g., students at one institution) based on a sample of that population (e.g., students on a course at that institution) using probability theory (Field, 2025). Inferential statistics also potentially allow researchers to make predictions - or inferences - and test hypotheses about the larger population based on the information gathered from a smaller sample, assuming the sample is representative and large enough.

Despite the importance of inferential statistics and the widespread availability of free statistical software, research in this field can rely on descriptive statistics (notably means or percentages), and use these to make either inaccurate causal claims, or claims about the larger population based purely on descriptives. For example, researchers may inaccurately report that an orientation or transition program resulted in (or *caused*) a higher mean sense of belonging score for students from working class backgrounds than students who did not take part in the program, and thus conclude that the program was effective. However, a descriptive account of the difference is insufficient, and does not tell us whether that difference is significant, in other words, is it meaningful in the real world? This is because the means from any two groups of people will always differ based on a number of underlying factors, which may or may not be relevant to the study.

Data collected from people often contains multiple sources of 'noise', including measurement errors, outliers, and confounding factors, which can obscure the true patterns underlying the data. There are different strategies for reducing this noise before data analysis, including data cleaning (e.g., identifying and removing outliers) and improving measurement reliability (e.g., using validated instruments). Next, inferential statistics allow researchers to determine whether any between-group differences are large enough to be important or *significant* (e.g., due to an intervention), or whether they reflect intra- or inter-individual variability that is unrelated to the research question.

The appropriate statistical test to use will be determined by the type of data that has been collected (see Table 1). Both nominal and ordinal data usually require nonparametric statistical tests, such as Mann-Whitney U test, Wilcoxon signed-rank test, Kruskal-Wallis test, Chi-squared test and Spearman's rank correlation coefficient. In reality, however, researchers usually convert agreement categories obtained from questionnaires using Likert scales (nominal data), into Interval/Ratio data by assigning each category of agreement with a numerical value (e.g., strongly disagree = 1, disagree = 2 etc.,). Nonparametric tests make fewer assumptions about the data than parametric tests, for example, data do not need to be normally distributed, nor have equal variances between groups. Nonparametric tests are less sensitive to skewed data and outliers than parametric tests, but this also renders them less powerful. In contrast, parametric tests can be used for Interval/Ratio data, and they are more sensitive and powerful than non-parametric tests. Basic common tests include t-tests, Analysis of Variance (ANOVA), and regression analysis. Although it is beyond the scope of this Editorial to provide guidance as to how to run and interpret these various tests, Field (2025) provides a comprehensive guide.

In summary, it is important that quantitative data in higher education learning and teaching research is analysed in a way that extends beyond merely describing its characteristics. Data should, where possible, be analysed using more sophisticated inferential statistics that enable the findings to be generalised beyond the sample, and potentially enable cause and effect to be determined. However, this necessarily requires good quality data to analyse in the first place. The issues outlined above in Principles 2 and 3 could be usefully summarised by the adage "measure twice, cut once" - researchers need to plan to collect good quality data that will support inferential statistical analysis *before* embarking on a research study: analysis, however sophisticated, cannot cannot fix poor data. As an example, a confounding variable that has not been measured during data collection cannot be accounted for during analysis by virtue of its absence. The key message here is that going beyond descriptive statistics increases the depth of insights we can gain from data and will improve our collective progress in understanding educational phenomena.

Table 1

Definitions of types of data and their advantages and disadvantages

Data Type	Definition	Examples	Advantages	Disadvantages
Nominal (Categorical)	Simple counts of observations that fall into mutually exclusive categories	Ethnic group, Preferred classroom seating (front, middle, back)	Simple and easy to analyze; Useful for identifying group differences	Cannot determine order or measure differences between categories
Ordinal	Data measured on a scale that enables ranking but does not indicate the magnitude of differences	Class rankings (1st, 2nd, 3rd), Students' course satisfaction (Dissatisfied, Neutral, Satisfied)	Allows for ranking and comparison; More informative than nominal data	Does not quantify the difference between ranks; Limited statistical analysis options
Interval	Measurements where differences between values are equal, but there is no true zero	Standardized test scores (IQ), Student grades (A, B, C, etc.)	Can compare differences between values meaningfully; Allows for advanced statistical analyses	No true zero point, meaning ratios are not meaningful (e.g., an IQ of 120 is not "twice as intelligent" as an IQ of 60)
Ratio	Measurements where differences between values are equal, and there is a true zero	Number of correct answers on a test, Number of absences in a semester	Allows for a full range of statistical analyses, including ratio comparisons	Requires precise measurement instruments

Recommendations for Researchers

Below is a summary of our key recommendations to researchers for carrying out high-impact quantitative higher education learning and teaching research, and disseminating it in journals, such as the Journal of University Teaching and Learning Practice. These are not intended to be prescriptive, but highlight key principles to support researchers to avoid common pitfalls as they take their first steps into this field.

Principle 1: Theorise well

- Does your research have a theoretical basis?
- Is the relevance of the theory justified in the manuscript?
- Do the measures align with the constructs of the chosen theory?
- Does your study have the potential to support or refute the theory?

Principle 2: Prioritise method robustness

- Does the design enable you to determine cause and effect, or does it merely describe relations between variables at a particular time point?
- Does your design control for potential confounding variables?
- Do you have an appropriate control group(s)?
- Where relevant, have you used a pre-existing questionnaire(s) with established reliability and validity?
- Is the sample drawn from the population to which you are generalizing your findings?
- Have you reported the sampling strategy, response rate, and fully described sample parameters?
- Are there any possible sampling biases that need to be acknowledged as limitations?
- Have you considered potential ethical issues and how to overcome them?

Principle 3: Analyse, do not merely describe

- Have you gone beyond descriptive analysis and applied appropriate inferential statistical analysis?
- Have you ensured that the underlying assumptions relating to your use of statistical tests have been met?

Conclusion

In this Editorial we have demonstrated that there is an urgent need for more high-quality quantitative research in higher education learning and teaching that provides a robust evidence base for improving students' experiences and academic outcomes. Drawing on limitations with research in this field, we have provided three Principles to help future researchers overcome these. The Principles highlight the need for theoretically-based, well-designed, and statistically analysed quantitative studies that enhance our understanding of learning and teaching. Addressing these issues will support researchers to move beyond providing narrow and descriptive accounts of teaching practices, and, instead, provide high-quality research that tells us how we might meaningfully make positive educational change for our students.

This is an exciting time for a field in which there are more issues than ever that need a high quality evidence-base to inform practice and policy. Notably, these include (but are not limited to) eliminating degree awarding gaps among unrepresented student groups (e.g., those from ethnic minorities); harnessing the potential of generative AI to support learning and teaching; improving student mental health and well-being; and developing effective assessment and feedback to enhance learning.

To address these issues, we need to prioritise interdisciplinary collaboration to ensure our research is of high quality, as well as the integration of diverse perspectives to ensure our research has wide reach and impact. At the institutional level, we call for leaders to invest in university teachers by providing them with resources to gain appropriate training in research methods and time to carry out research. This is critical if we are to raise the status of pedagogic research that underpins SoTL and contribute to a more evidence-based and equitable future for our students.

Acknowledgements

The authors disclose that they have no actual or perceived conflicts of interest nor have they received any funding for this manuscript beyond resourcing for academic time at their respective universities. The authors used generative artificial intelligence minimally to support the initial ideation of this article, and to help with proof-reading for clarity and conciseness. The authors list the following CRediT contributions: **Taylor:** Conceptualisation, Project Administration and Supervision, Visualisation, Writing – Original Draft, Writing – Review & Editing. **Dai**: Conceptualisation, Writing – Original Draft, Writing - Review & Editing. **Playfoot:** Conceptualisation, Writing - Original Draft, Writing - Review & Editing. **Rose:** Conceptualisation, Writing – Original Draft, Writing – Review & Editing. **Ng:** Conceptualisation, Writing – Original Draft, Visualisation. **Purvis, Alison:** Writing – Review & Editing. **Crawford, Joseph:** Conceptualisation, Writing – Review & Editing.

References

Ali, A., Crawford, J., Cejnar, L., Harman, K., & Sim, K. N. (2021). What student evaluations are not: Scholarship of Teaching and Learning using student evaluations. *Journal of University Teaching and Learning Practice*, *18*(8). https://doi.org/10.53761/1.18.8.1

Ashwin, P. (2012). How often are theories developed through empirical research into higher education? *Studies in Higher Education*, *37*(8), 941–955. https://doi.org/10.1080/03075079.2011.557426

Berliner, D.C., (2002). Comment: educational research: the hardest science of all, *Educational Researcher, 31*(8), 18–20. https://www.jstor.org/stable/3594389

Boyer, E. (1990). Scholarship reconsidered: Priorities of the professoriate. *The Carnegie Foundation for the Advancement of Teaching*. Princeton University Press.

Bryman, A. (1984). The debate about quantitative and qualitative research: A question of method or epistemology? *British Journal of Sociology*, *35*(1), 75–92. https://doi.org/10.2307/590553

Bull, S., Cooper, A., Laidlaw, A., Milne, L., & Parr, S. (2024). 'You certainly don't get promoted for just teaching': experiences of education-focused academics in research-intensive universities. *Studies in Higher Education*, *50*(2), 239–255. https://doi.org/10.1080/03075079.2024.2333946

Cage, E., Jones, E., Ryan, G., Hughes, G., & Spanner, L. (2021). Student mental health and transitions into, through and out of university: student and staff perspectives. *Journal of Further and Higher Education*, *45*(8), 1076–1089. https://doi.org/10.1080/0309877X.2021.1875203

Canning, J., & Masika, R. (2020). The scholarship of teaching and learning (SoTL): the thorn in the flesh of educational research. *Studies in Higher Education*, *47*(6), 1084–1096. https://doi.org/10.1080/03075079.2020.1836485 6.

Cotton, D. R. E., Miller, W., & Kneale, P. (2018). The Cinderella of academia: Is higher education pedagogic research undervalued in UK Research Assessment? *Studies in Higher Education, 43,* 1625–1636. https://doi.org/10.1080/03075079.2016.1276549

Creswell, J. W., & Creswell, J. D. (2022). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches,* (6th ed). Sage Publications.

De Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory.* Guilford Press.

Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and Perspectives, 38*(1)*,* 105–123. https://search.informit.org/doi/abs/10.3316/INFORMIT.491551710186460

Ejnavarzala, H. (2019). Epistemology–ontology relations in social research: A review. *Sociological Bulletin*, *68*(1), 94–104. https://doi.org/10.1177/00380229188193

ERA (2018). *Evaluation Handbook*. Canberra: Commonwealth of Australia, Australian Government Australian Research Council. https://www.arc.gov.au/ei-key-documents

Evans, C., Muijs, D. & Tomlinson, M. (2015). *Engaged student learning: High impact strategies to enhance student achievement.* Higher Education Academy. Available at: https://www.advance-he.ac.uk/knowledge-hub/engaged-student-learning-high-impact-strategies-enhance-student-achievement

Evans, C., Kandiko Howson, C., Forsythe, A., & Edwards, C. (2021). What constitutes high quality higher education pedagogical research? *Assessment and Evaluation in Higher Education*, *46*(4), 525–546. https://doi.org/10.1080/02602938.2020.1790500

Field, A. P. (2025). *Discovering Statistics Using R and RStudio (2nd Ed)*. SAGE Publications.

Hartas, D. (Ed.). (2015). *Educational Research and Inquiry: Qualitative and Quantitative Approaches*. Bloomsbury Publishing.

Hjørland, B. (2005). Empiricism, rationalism and positivism in library and information science. *Journal of Documentation, 61*(1)*,* 130–155. https://doi.org/10.1108/00220410510578050

How, Z. J. (2020). A systematic review of scholarship of teaching and learning research in higher education institutes from 2014–2019. In S. Tan and S. H. Chen (Eds.), *Transforming Teaching and Learning in Higher Education.* Springer. https://doi.org/10.1007/978-981-15-4980-9_2

Howard, C., & Brady, M. (2015). Teaching social research methods after the critical turn: challenges and benefits of a constructivist pedagogy. *International Journal of Social Research Methodology*, *18*(5), 511–525. https://doi.org/10.1080/13645579.2015.1062625

Jabrayilov, R., Emons, W. H. M., & Sijtsma, K. (2016). Comparison of Classical Test Theory and Item Response Theory in individual change assessment. *Applied Psychological Measurement, 40*(8), 559–572. https://doi.org/10.1177/0146621616664046

Kanaki, K., & Kalogiannakis, M. (2023). Sample design challenges: an educational research paradigm. *International Journal of Technology Enhanced Learning, 15*(3), 266–285. https://doi.org/10.1504/ijtel.2023.10055808

McAvoy, P., Hunt, T., Culbertson, M. J., McCleary, K. S., DeMeuse, R. J., & Hess, D. E. (2021). Measuring student discussion engagement in the college classroom: a scale validation study. *Studies in Higher Education*, *47*(8), 1761–1775. https://doi.org/10.1080/03075079.2021.1960302

Macfarlane, B. (2011). Prizes, pedagogic research and teaching professors: Lowering the status of teaching and learning through bifurcation. *Teaching in Higher Education, 16*(1)*,* 127–30. https://doi.org/10.1080/13562517.2011.530756

Pownall, M., Birtill, P., & Harris, R. (2025). Towards conceptual clarity in pedagogical research: the case of the 'imposter phenomenon' jingle/jangle fallacy. *Higher Education Research and Development*, https://doi.org/10.1080/07294360.2025.2470821

Purvis, A. J. & Crawford, J. (2024). Ethical standards in educational research and publication. *Journal of University Teaching and Learning Practice*, *21*(9). https://doi.org/10.53761/hqnqr710

Purvis, A., Nicholas, V., & Tai, J. (2024). What's your problem? Writing effective research questions for quality publications. *Journal of University Teaching and Learning Practice*, *21*(10). https://doi.org/10.53761/j64xa573

Raykov, T., & Marcoulides, G. A. (2016). On the relationship between Classical Test Theory and Item Response Theory: From one to the other and back. *Educational and Psychological Measurement, 76*(2), 325–338. https://doi.org/10.1177/0013164415576958

REF (2019). *Guidance on Submissions.2019/01.* Research Excellence Framework. Available at: https://www.ref.ac.uk/publications

Rowland, S. L., & Myatt, P. M. (2014). Getting started in the scholarship of teaching and learning: A "how to" guide for science academics. *Biochemistry and Molecular Biology Education, 42*(1)*,* 6–14. https://doi.org/10.1002/bmb.20748

Sharkness, J., & DeAngelo, L. (2011). Measuring student involvement: A comparison of classical test theory and item response theory in the construction of scales from student surveys. *Research in Higher Education, 52*, 480–507. https://doi.org/10.1007/s11162-010-9202-3

Taber, K.S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*, 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Thiem, J., Preetz, R., & Haberstroh, S. (2023). How research-based learning affects students' self-rated research competences: evidence from a longitudinal study across disciplines. *Studies in Higher Education*, *48*(7), 1039–1051. https://doi.org/10.1080/03075079.2023.2181326

Tight, M. (2017). Tracking the scholarship of teaching and learning. *Policy Reviews in Higher Education*, *2*(1), 61–78. https://doi.org/10.1080/23322969.2017.1390690

Toland, M. D. (2014). Practical guide to conducting an Item Response Theory analysis. The *Journal of Early Adolescence, 34*(1), 120–151. https://doi.org/10.1177/0272431613511332

Trigwell, K. (2013). Evidence of the impact of scholarship of teaching and learning purposes. *Teaching and Learning Inquiry*, *1*(1), 95–105. https://www.jstor.org/stable/10.2979/teachlearninqu.1.1.95

Vignoli, M., Dosi, C., & Balboni, B. (2023). Design thinking mindset: scale development and validation. *Studies in Higher Education*, *48*(6), 926–940. https://doi.org/10.108

Waller, K. L., & Prosser, M. (2023). The rapidly changing teaching and research landscape: The future of SoTL and the teaching-research nexus. In K. Coleman et al. (eds.), *The*

Educational Turn: Rethinking the Scholarship of Teaching and Learning in Higher Education (pp. 27–41). Springer Nature Singapore.

Xia, Q., Chiu, T. K., & Chai, C. S. (2023). The moderating effects of gender and need satisfaction on self-regulated learning through Artificial Intelligence (AI). *Education and Information Technologies*, *28*(7), 8691–8713. https://doi.org/10.1007/s10639-022-11547-x

Zhao, L., Peng, J., Yang, X., Yan, W., Ke, S., Batool, K., ... & Lee, K. (2024). Effects of academic dishonesty policy reminder on university students' exam cheating–a double-blind randomized controlled experimental field study. *Studies in Higher Education, 49*(4), 592–608. https://doi.org/10.1080/03075079.2023.2246073