

## Using quantitative data

Michael Jopling

### Introduction

‘98% of statistics are made up’ (Anon)

If the thought of working with quantitative data frightens you or leaves you cold, this chapter is designed to make you feel more confident. The intention is to demystify quantitative research and give you the basic information you need to understand how it works. While quantitative research is essentially about measurement and using numbers to generate breadth and generalizability (Basit, 2010), the chapter does not call for any quantitative or statistical expertise and contains nothing more than basic mathematics. Its focus is on outlining the logic and some of the methods and key terms that lie behind quantitative approaches and the use of quantitative data, with examples taken from real research.

It is important to emphasize that I approach this subject not as a statistician, but as someone who has come to recognize the power and value of incorporating quantitative data into educational investigations. Anyone seeking more detailed consideration of the issues discussed, particularly in relation to data analysis, is referred to the methodological texts referred to throughout. They all offer invaluable information and in-depth explanation.

It is still common for researchers working in the education field to regard quantitative/qualitative as a methodological divide, in which they have to choose a side, rather than merely as types of data. This chapter rejects this division and takes as its starting point the notion that researchers in this area need to understand how *both* quantitative *and* qualitative data are collected, analysed and interpreted. Here are three reasons why readers who are nervous of numbers should nevertheless consider working with quantitative data:

- Having some understanding of quantitative data will improve your ability to assess investigations which use quantitative data and approaches, particularly if you are new to research.

- Being able to incorporate quantitative data into your own research may strengthen its impact, particularly if you are interested in influencing aspects of education practice or policy. Although they often use them carelessly and unreflectively, policymakers like numbers.
- Having an improved understanding of quantitative data and research will put you in a better position to evaluate or conduct mixed-methods investigations.

### **What is quantitative research? What are quantitative data?**

The application of quantitative methods in social-science research goes back to interest in the 19<sup>th</sup> century in applying forms of experimental investigation, developed in the natural sciences. This sought to understand the processes of cause and effect through manipulation and controlled testing. However, as it became clear that many of the issues explored in social sciences, such as education, do not lend themselves to experimental approaches, researchers began from the 1950s to apply them to what became known as ‘quasi-experimental’ and ‘non-experimental’ situations. Provocatively, Berliner (2002:19) has described what social scientists and particularly educational researchers do as the ‘hardest to do science [...] because humans in schools are embedded in complex and changing networks of social interaction’. It is engaging with this social complexity that makes the application of experimental research to education so challenging (Hadfield and Jopling, 2018). It is partly as a consequence of the need to address this complexity and the limitations of the persistent quantitative/qualitative divide, manifested in the so-called ‘paradigm wars’, that mixed-methods research design (Tashakkori and Teddlie, 1998) gained ground in the 1990s.

There are also many definitions of quantitative research. Aliaga and Gunderson (2002, cited in Muijs, 2011: 1) describe it as ‘explaining phenomena by collecting numerical data that are analysed using mathematically based methods (in particular statistics)’. Punch and Oancea (2014:264) state that quantitative research essentially does three things: it conceptualizes reality in terms of variables, measures these variables, and studies relationships between these variables. But what are ‘variables’?

### *Variables: Understanding cause and effect*

Quantitative research is essentially concerned with cause and effect. This is expressed in terms of variables. A *variable* is simply a construct or characteristic which a researcher is interested in observing, measuring or manipulating. Age, ethnicity and experience are all possible variables.

Quantitative research distinguishes between independent and dependent variables. The *independent variable* is also known as the input, treatment or predictor variable. It is often a stimulus, intervention or treatment, like a new teaching approach, that causes a specific outcome. A *dependent variable* is the outcome or criterion variable and can be regarded as the effect or consequence of the use of the independent variable.

‘Control’ is a central concept in quantitative research design. It refers to attempts to minimize the effects of variables, other than that of the independent variable. As researchers, we want to control these *extraneous variables* to prevent them affecting the results of the experiment. In experimental research therefore, participants are assigned randomly either to the group which has the intervention or the group which does not (the ‘control’ group), so that there are no systematic differences between them. Or they are matched purposefully on a number of characteristics, so that the two groups are as close as possible when the experiment begins. In quasi-experimental research, which is research where such randomization or matching is not possible, the researcher uses statistical methods to control (or remove) the effects of other variables which may influence their results.

If variables are not controlled, or cannot be controlled, they become confounding variables. A *confounding variable* is a kind of extraneous variable, separate from independent and dependent variables, which may influence those variables and the relationships between them. An *intervening variable* (also known as the control or mediating variable, or covariate) is another kind of extraneous variable, one which helps explain the relationship between the dependent and independent variables. For example, research has found a relation between an individual’s level of education and their level of income, but this is not a direct cause-and-effect relationship – intervening variables, such as type of occupation, also play a part.

Here is a fuller example of the above notions. If we were interested in comparing the resilience levels of girls and boys, that is, their respective capacity to cope with adversity,

resilience (however measured) would be the dependent variable and gender would be the independent variable. Extraneous variables could include the time of day when the research was conducted or the participants' ages. We would wish to take steps to control these extraneous variables, for example by ensuring that participants were all tested at the same time on the same day and that they were all within a certain age range. If this was part of research into the effectiveness of an intervention designed to increase their resilience, we would want to measure the young people's resilience levels before and after the intervention.

[BOX BEGINS]

From your reading of the chapter so far, as well as drawing on your own ideas:

1. What kinds of research issues might you want to explore where quantitative data could play a part?
2. How might you determine the aims, schedules and even the costs involved in such a project?
3. What kinds of data would you need to collect and how might you translate your data into variables?

[BOX ENDS]

### **Three types of quantitative design**

As you can see from the chapter so far, there are different ways to categorize quantitative research design, each referring to the extent to which they are experimental. The categories usually used are experiments, quasi-experiments and non-experiments.

#### *Experiments*

Experiment is used here as a precise technical term, defined as 'a test under controlled conditions that is made to demonstrate a known truth or examine the validity of a hypothesis' (Muijs, 2011:11). Scott and Usher (2011:61) state that 'the experimental researcher attempts to explicate causal relationships between phenomena by intervening in the natural setting and controlling the relevant variables'. True experiments also involve randomizing participation and controlling variables.

The concepts of internal and external validity (Campbell and Stanley, 1963) are relevant in this context. *Internal validity* refers to the extent to which researchers can be sure that the effects they ascribe to the interventions under examination are not caused by other factors. In education contexts, such factors could include participants' maturation or socio-economic status, alongside broader issues, such as sampling and the reliability of the test itself. *External validity* refers to the extent to which findings can be generalized to larger populations.

Until recently, much experimental research in education was really quasi-experimental. This changed once randomized controlled trials (RCTs) started to be promoted and funded by governments in England and elsewhere as a 'gold standard'. Imported from medical research, RCTs involve randomization and pre- and post-testing (Connolly *et al.*, 2017). They are typically large-scale and expensive. Although you would not always think so from the evangelical terms in which RCTs are often described, a number of issues and limitations have been identified. Goldacre (2013:13), an advocate of applying RCTs to educational issues, was clear that 'randomised trials are very good at showing *that* something works; they're not always so helpful for understanding *why* it worked' (italics in original), a limitation which is more likely to be addressed by using qualitative methods. Morrison (2001) and Wrigley (2018) have criticized RCTs for taking a restricted view of causality and neglecting the contexts and complexity of the phenomena being researched.

One of the most commonly-heard criticisms of using RCTs in education research is that it is unethical to exclude students through randomization. Goldacre's (2013) fairly robust response was that RCTs are about deciding whether an intervention works before implementing it, and that successful interventions can and should subsequently be offered to 'non-treated' participants. Hanley *et al.* (2016) argue that viewing RCTs as the gold standard has obscured the value of integrating the use of other, more implementation-specific measures. All of this suggests that you need to be aware of the ethical and pragmatic issues involved when planning an RCT and should never assume that social research of any kind is value-free.

The Educational Endowment Foundation (EEF), established in 2010, has created a Teaching and Learning Toolkit which summarizes the international evidence on teaching in the compulsory phases and commissions projects, which are evaluated using RCTs (Higgins *et al.*, 2016). This is a good source of examples of experimental research and, at the time of

writing, featured reports on 90 intervention projects. The most highly rated in terms of effectiveness was one which used an approach called ‘Self-Regulated Strategy Development’ to help struggling writers in the final year of primary school and the first year of secondary school in England. The evaluation (Torgerson and Torgerson, 2014) involved 23 West Yorkshire primary schools, randomly allocated to the intervention. Schools assigned to the control group were offered the intervention training at the end of the trial. The evaluation found that the approach had a strong positive effect on the writing outcomes of low-attaining students, particularly those from low-income families and therefore eligible for free school meals.

### *Quasi-experiments*

As already indicated, quasi-experimental research is used when randomization is not possible and therefore extraneous variables cannot be controlled physically. Quasi-experimental research compares groups and tries to assess and extract the variance between them through statistical procedures. The advantages are that you can investigate and compare naturally occurring groups, for example within a school or between different schools, which are already representative of aspects of the real world, although this requires you to think carefully about issues of validity and generalizability. One major limitation is related to the possibility of selection bias, which is why it is crucial for anyone using quasi-experiments to justify sampling procedures in detail in the study’s research design.

In our study of practice leadership in the early years (Hadfield *et al.*, 2015), we used structured observations of interactions between adults and children in early-years settings to evaluate practitioners’ impact on children’s learning and development. Adopting a mixed-methods approach, we undertook multiple sets of two types of observation: rating of the pre-school environment and observations of interactions between adults and children. With these we assessed the quality of provision and the impact of improvement interventions over three years. We used correlation analysis, which quantifies how closely variables are related, to investigate impact and to construct from the data ‘improvement trajectories’ for the 25 case-study settings, which were not randomized. The challenges involved included ensuring observational reliability in a large research team and maintaining participants’ engagement and involvement in the investigation. We tried to address the latter by sharing initial analysis of the quantitative data with participants to stimulate their reflections and explanations in the

qualitative elements of the research, for example in relation to changes in the kinds of interactions between adults and children that had been observed.

### *Non-experiments*

If you are a relatively inexperienced researcher, survey or ‘correlational’ research is likely to be the most common form of quantitative investigation you undertake. Such research is described as non-experimental, because there is no randomization or experimental manipulation, and therefore no attempt to identify cause-and-effect relationships. Instead, it examines correlations between variables - how and how much they are related to each other (covered in a little more detail in the next section). Surveys are particularly good for capturing the attitudes or characteristics of a population, for example individuals’ beliefs, opinions or behaviours. However, surveys often do this without acknowledging the messiness and complexity of social reality, represented not least by the fact that people interpret questions in different ways. This suggests that findings from surveys need to be very clearly described and heavily caveated.

Questionnaire surveys can be conducted face to face or by telephone, but here I focus predominantly on those that are completed in writing, either online or in hard copy. Essentially, there are two types: cross-sectional and longitudinal (Creswell, 2018). *Cross-sectional* questionnaire surveys give you a snapshot of views at a single point in time; *longitudinal* surveys are repeated and enable you to see developing trends. Both kinds are relatively quick and inexpensive ways to get a breadth of views. Online questionnaires, which are increasingly popular, have advantages in terms of accessibility and sample size, but also create challenges in authenticating respondents and ensuring that samples are representative. There are also large, regular, national surveys, such as the Millennium Cohort Study and British Social Attitudes survey, that can be mined to inform your research. Surveys of all kinds need careful planning and piloting to ensure that key factors, such as form, content, language, sequencing, and (particularly) length, are appropriate. See Oppenheim (1992) or Hartas’ (2010:266) checklist of common problems for guidance.

[BOX BEGINS]

### **Design**

1. What benefits do you see in undertaking experimental, quasi-experimental or non-experimental research?

2. What challenges are associated with each and how might these be addressed?
3. Which approach might be most appropriate for quantitative research you are intending to carry out yourself?

[BOX ENDS]

## **Analysing quantitative data**

As indicated at the beginning, there is not scope here to offer more than a brief introduction to some of the issues involved in quantitative data analysis. Therefore, this section can be regarded as a kind of extended glossary, which introduces some of the key terms you will need in order to understand other people's quantitative research and to begin thinking about conducting your own.

In order to collect quantitative data, as in all kinds of research, you need to be clear about what you want to find out. Your methods and your analysis will be derived from your hypotheses or research questions, as part of your research design. One of the key issues to understand is the notion of measurement itself. Punch and Oancea (2014:287) define it as 'the process of using numbers to link concepts to indicators, where there is a continuum involved', that is, assigning numerical values to ideas. They identify six main steps in constructing a measuring instrument, which include defining what you are planning to measure, selecting a measuring technique (such as a Likert scale – see below), and generating items – statements or questions – to use. You also need to decide whether to construct a new instrument, adopt an existing one, or use a combination of the two. In the practice leadership study (Hadfield *et al.*, 2015), we used validated rating scales for evaluating the pre-school setting environments and devised a new observation tool, based on findings from relevant research, to assess the interactions between adults and children. Using these tools in combination enabled us to create the quantitative baseline and progress measures for the case studies we developed.

### *Levels of measurement*

In quantitative research, there are four levels of measurement: nominal, ordinal, interval, and ratio. It is important to understand these levels, as they determine the kinds of analysis you can apply to your data.



*Nominal scales* assign numerical values arbitrarily. Gender is a good example of this: female may be assigned the value '1', and male the value '2'. These are mutually exclusive categories and have no logical order or ranking in relation to each other. Being labelled '2' does not make an individual greater, better or worse than being labelled '1', and *vice versa*.

In *ordinal scales*, characteristics are assigned numerical values which allow researchers to rank survey responses on a continuum, for example in relation to their frequency or importance. Political orientation – left, right and points between – can be measured on an ordinal scale. A Likert scale, commonly used in questionnaire surveys, is probably the most common example of an ordinal scale, in which respondents are asked to select from a range of responses, such as 'Strongly agree', 'Agree', 'Neither agree nor disagree', 'Disagree' and 'Strongly disagree'. The intervals between these responses have a logical order, but they are not equal, and, like nominal scales, they are mutually exclusive categories.

*Interval scales* have equal intervals between points, like a ruler. As with ordinal scales, these points are mutually exclusive and logically ordered, but they differ in that they are continuous and the distance between them is assumed to be equal. However, they do not have a true zero point corresponding to the absence of the quality being measured. Time of day on a 12-hour clock is an example of an interval scale, as are numerical scales of measurement for opinions or attitudes, for instance running from 1 to 10.

Like interval scales, *ratio scales* have equal distances between their units. However, they do have an absolute zero point. Weight and height are examples of ratio variables measured with ratio scales, as the difference between 45kg and 50kg, for example, is the same as the difference between 55kg and 60kg. 24-hour time, where midnight = 0, is another ratio scale.

Interval and ratio scales create *parametric* (or continuous) data. Nominal and ordinal scales create *non-parametric* (categorical or non-continuous) data. This is important, because as we move through these levels from nominal to ratio, the procedures we can use in statistical analysis become less restricted. Parametric data can be analysed using a wide range of techniques, including regression and factor analysis (see below). Non-parametric data can be analysed only by using a narrower set of analytical tools. Parametric data are regarded as more robust than non-parametric data. Hartas (2010:350) has another good checklist you can

refer to in order to establish what kinds of tests can be used with what kinds of variables and data.

Most quantitative data analysis is undertaken with statistical software, the most commonly used of which is SPSS. A number of excellent guides to this exist, relating both to social science in general (Field, 2018) and to education more specifically (Muijs, 2011). Microsoft Excel can also be used for many analytical tasks. However, before you begin, you need to know what kind of analysis is appropriate – for your data, for answering your research questions or for testing your hypothesis. The two main types, descriptive analysis and inferential analysis, are covered below.

### **Descriptive statistics**

Analysis usually starts with descriptive statistics, which summarize your data. Nominal and ordinal data are described in terms of frequencies, percentages or cross-tabulations. Interval and ratio data allow centrality and dispersion to be identified – covered below. Such statistics are an essential part of quantitative analysis and can take you a long way, particularly in mixed-methods research. Indeed, depending on your research design, a descriptive analysis may be all you need, especially in a small-scale or pilot project, and the analysis can be done using software, such as SPSS.

#### *Central tendency: Mean, mode and median*

Central tendency is a way of describing what is typical in a dataset. It has three main measures: mean, median and mode. The *mean* is a hypothetical value, colloquially referred to as the ‘average’, calculated by dividing the total of all the values in a dataset by the number of values in that set. While the mean is often calculated for ordinal variables, it makes most sense for continuous data (interval and ratio variables). The *mode* is simply the value which occurs most frequently in a dataset and therefore applies to nominal, ordinal, interval and ratio variables. The *median* is the middle category in a distribution of values, arranged by ordering them from high to low. For example, in a dataset of ages which runs 28, 27, 25, 25, 23, 22, 20, the median age is 25 (the mode is also 25; the mean is 24).

### *Dispersion: Range and variance (standard deviation)*

Dispersion indicates how the values are spread around the central tendency. The two most common measures of dispersion are range and variance. The *range* is calculated simply by subtracting the lowest from the highest value in a distribution of values from an interval or ratio variable. For example, in the dataset for which the median was calculated above, the range is 8 (28-20). *Variance* is the measure of the spread and indicates how close or how far the observed scores or values are from the mean. A large variance suggests that the mean is not a good representation of the data. A small variance indicates that the mean is a good representation, as the data cluster around it. In this connection, you will often find reference to the *standard deviation* (SD). SD is the square root of the variance and is considered to be a more accurate description of dispersion than the range between extreme scores (known as *outliers*). SD is an important concept to understand, because parametric data are assumed to come from a population which is normally distributed. This is an idealized or standard distribution with a 'bell curve' shape, which assumes a mean of 0 and a standard deviation of 1. SD is included in SPSS outputs and data tables.

## **Inferential statistics**

Before outlining some of the methods that you can employ in inferential statistics, it is crucial to introduce two key ideas which you will almost always come across when you read research reports with quantitative findings: statistical significance and effect size.

### *Statistical significance*

Statistical significance simply refers to the probability that a relationship or difference between variables occurred by chance (Creswell, 2018). Conventionally, the threshold or critical value for the significance level is set at .05 (5 in 100 or 1 in 20). A high level of statistical significance would be .01 (1 in 100). It is usually represented as a  $p <$  value. Statistically significant results are more likely to be found with large sample sizes. In recent years, statistical significance has been increasingly criticized for shortcomings, such as its arbitrary cut-off points, and more emphasis has been placed on measuring effect sizes (Muijs, 2011).

### *Effect size*

Effect sizes have become increasingly prominent in education research since the publication of Hattie's (2009), *Visible Learning*, which is a synthesis of over 800 meta-analyses of education research. Hattie used the effect size of quantitative studies to aggregate information about a range of interventions and to rate their effectiveness. Effect size is the size of the difference between groups or of the relationships between variables. Whereas statistical significance only tells you the likelihood of an outcome occurring by chance, effect size helps to identify the strength of the conclusions of a research study. While there are different ways to calculate effect sizes, the output value ranges between 1 and -1. Usually an effect size of greater than 0.5 is considered a strong effect; between 0.3 and 0.5 is a moderate effect; 0.1-0.3 is a modest effect; and 0-0.1 is a weak effect. As you would imagine, negative effect sizes identify negative effects.

Alongside estimations of the strength of the evidence associated with an educational intervention or teaching approach, and of its cost, the EEF Teaching and Learning Toolkit, mentioned earlier, translates effect sizes into additional months' progress in learning that students can make as a result of that particular teaching approach being used. In the 'Self-Regulated Strategy Development' project, used as an example above, the overall effect size of the intervention was 0.74, which was translated in the Toolkit into an additional nine months' progress for the school students involved. The effect size was higher for those eligible for free school meals. While the Toolkit offers useful summary guidance for practitioners, the months' progress figure has been criticized for collating findings from very different studies and contexts into a single, potentially misleading, headline progress figure (Wrigley, 2018).

### *Tests of difference and variance*

The *chi-square* test examines association or, more usually, difference. It measures the variance between a result that is expected statistically and the actual result, to determine whether it is significant. The *t-test* determines if there is a significant difference between the means of two randomly selected groups with a normal distribution of scores. You will also often find reference to ANOVA – analysis of variance. Whereas t-tests focus on two groups, two variables or two timepoints, ANOVA is used to compare and examine differences between more than two groups, all of which have been randomly selected and have normal distribution. *Equivalent tests* have been created to use with non-parametric data. Available in software like SPSS, the Mann-Whitney U and Wilcoxon tests are alternatives to the t-test; the

Kruskal-Wallis and Friedman tests are the non-parametric equivalents to ANOVA.

### *Correlations*

Correlations are used to test hypotheses about the relationships between variables. They measure the extent to which change in the value of one variable is accompanied by a change in the value of another variable. They are calculated as a *correlation coefficient*, expressed as a score between +1 and -1. +1 is a perfect positive relationship; -1 is a perfect negative relationship; and 0 is no relationship. The closer the score is to +1, the stronger the relationship. For example, in our research into effective university teaching (Allen *et al.*, 2009), we found that the strongest correlation in a 32-item questionnaire, with a correlation coefficient of 0.829, was between lecturers' being enthusiastic about learning and their respecting of students' opinions. It cannot be emphasized enough that while correlation identifies a relationship between variables, even a very strong correlation never implies causation.

### *Factor analysis*

Factor analysis uses correlation to reduce the number of variables in a dataset. By factor-analysing the correlations between variables, it creates a smaller number of derived variables, referred to as *factors*. For example, test results from a range of subjects, such as physics, chemistry, sociology and psychology, can be reduced to two groups: sciences and social sciences (depending on where you see the dividing line) in order to simplify analysis.

### *Regression analysis*

Regression analysis, or more accurately *multiple linear regression*, is used when you want to study the relationship between a dependent variable and a number of independent variables. Punch and Oancea (2014:279) use the example of school achievement as the dependent variable, in relation to which the effect of independent variables, such as socio-economic status, motivation and levels of homework, can be investigated using regression analysis.

### *Presenting quantitative data*

Although they invariably contain examples, most research-methods guides devote much more space to the analysis of data than they do to how to present those data and their analysis. While journal articles generally contain data tables, it is still all too common to see standard Excel or SPSS charts exported without modification into student assignments, research

reports or conference presentations. Graphs and charts can be a very effective way of conveying complex findings to an audience, but you need to ensure that they are not overly complicated, that you can explain them either verbally or in writing, and that they are appropriate for the audience towards which they are targeted. Testing graphs with colleagues, or getting input from information designers, will help.

[BOX BEGINS]

### **Analysis and presentation**

1. How confident are you about using descriptive analysis in your quantitative research?  
What planning do you need to do?
2. If you are thinking about using a more sophisticated approach, such as inferential statistics, what kind of analysis do you need to do and, more importantly, who are you going to ask to help you?
3. How are you going to present your quantitative data and its analysis?

[BOX ENDS]

### **Conclusion**

Scott and Usher (2011) conclude their overview of experimental and correlational research with two points that are worth reiterating here. The first is that the open systems in which we operate in social-science research make accurate prediction impossible. The second is that attempts to model school processes mathematically are necessarily reductive of the complexity of these systems and fail to incorporate the effects of individuals' intentions, beliefs and attitudes. This is not to delegitimize the role of quantitative data in areas like education, but to acknowledge the importance of recognizing that determining 'what works' in quantitative terms is only part of the story. Equally, however, if researchers and practitioners, open to the idea of evidence-informed practice, are reluctant or unwilling to engage with the procedures or findings of quantitative research, they are closing themselves off from evidence, approaches and ideas which may be of benefit. Researchers, therefore, need to be able to draw on multiple research methods, in which a knowledge of quantitative data is necessary, but not sufficient. As Berliner (2002:20) emphasized: 'A single method is not what the government should be promoting for educational researchers. It would do better by promoting argument, discourse, and discussion.'

To conclude, do not be afraid of quantitative research or quantitative data. If you are drawing on the investigations and findings of others to inform your learning or your teaching, an understanding of the basic tenets of quantitative research design will enable you to evaluate the value of the reports you encounter more effectively. If, however, you are doing your own research, start with your research issue, questions or hypothesis, and think about what kinds of data you need to collect. Descriptive analysis can take you a long way, but work with more expert colleagues if you want to go further and move into inferential analysis. Finally, think carefully about how you present your data. A bad graph may lose you your audience; a good graph can convey your findings powerfully, and more quickly than a lengthy description.

### **Recommended reading**

Punch and Oancea's (2014) *Introduction to Research Methods in Education* has three chapters on quantitative research and is an excellent overview. Muijs's (2011) *Doing Quantitative Research in Education with SPSS* is a relatively brief guide and a good introduction to inferential analysis. Field's (2018) *Discovering Statistics* is the canonical text, which includes all the explanations (and the mathematics) you will need and the kind of user-friendly design you will either love or hate.

### **References**

Aliaga, M. and Gunderson, B. (2002) *Interactive Statistics*. Second Edition. Saddle River, NJ: Prentice-Hall.

Allan, J., Clarke, K. and Jopling, M. (2009) Effective teaching in higher education: Perceptions of first year undergraduates. *International Journal of Teaching and Learning in Higher Education*, 21(3), 362-372.

Basit, T.N. (2010) *Conducting Research in Educational Contexts*. London: Continuum.

Berliner, D.C. (2002) Comment: Educational research: The hardest science of all. *Educational Researcher*, 31(8), 18-20.

Campbell, D.T. and Stanley, J.C. (1963) Experimental and quasi-experimental designs for research. In: N.L. Gage, *Handbook of Research on Teaching*. Chicago, IL: Rand McNally.

Connolly, P., Biggart, A., Miller, S., O'Hare L. and Thurston, A. (2017) *Using Randomised Controlled Trials in Education*. London: Sage.

Creswell, J. (2018) *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. Sixth edition. London: Pearson.

Field, A. (2018) *Discovering Statistics Using IBM SPSS Statistics*. Fifth Edition. London: Sage.

Goldacre, B. (2013) *Building Evidence into Education*. London: Department for Education. Retrieved from <http://media.education.gov.uk/assets/files/pdf/b/ben%20goldacre%20paper.pdf>, 9 July 2018.

Hadfield, M. and Jopling, M. (2018) Case study as a means of evaluating the impact of early year leaders: Steps paths and routes. *Evaluation and Program Planning*, 67, 167-176.

Hadfield M., Jopling, M. and Needham, M. (2015) *Practice Leadership in the Early Years: Becoming, Being and Developing as a Leader*. Maidenhead: Open University Press.

Hanley, P., Chambers, B. and Haslam, J. (2016) Reassessing RCTs as the 'gold standard': Synergy not separatism in evaluation designs. *International Journal of Research and Methods in Education*, 39(3), 287-298.

Hartas, D (Ed.) (2010) *Educational Research and Inquiry: Qualitative and Quantitative Approaches*. London: Continuum.



Hattie, J. (2009) *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. London: Routledge.

Higgins, S., Katsipataki, M., Villanueva-Aguilera, A.B., Coleman, R., Henderson, P., Major, L.E., Coe, R. and Mason, D. (2016) *The Sutton Trust-Education Endowment Foundation Teaching and Learning Toolkit*. London: Education Endowment Foundation.

Morrison, K. (2001) Randomised controlled trials for evidence-based education: Some problems in judging 'what works'. *Evaluation & Research in Education*, 15(2), 69-83, DOI: 10.1080/09500790108666984.

Muijs, D. (2011) *Doing Quantitative Research in Education with SPSS*. Second Edition. London: Sage.

Oppenheim, A.N. (1992) *Questionnaire Design, Interviewing and Attitude Measurement*. Second edition. Reprint. London: Continuum, 1998.

Punch, K.R. and Oancea, A. (2014) *Introduction to Research Methods in Education*. Second Edition. London: Sage.

Scott, D. and Usher, R. (2011) *Researching Education: Data, Methods and Theory in Educational Enquiry*. Second edition. London: Continuum.

Tashkkori, A., and Teddlie, C. (1998). *Mixed Methodology: Combining Qualitative and Quantitative Approaches*. Thousand Oaks: Sage.

Torgerson, D. and Torgerson, C. (2014) *Improving Writing Quality: Evaluation Report and Executive Summary*. London: Education Endowment Foundation.

Wrigley, T. (2018) The power of 'evidence': Reliable science or a set of blunt tools? *British Educational Research Journal*, 44(3), 359-376.