PREPRINT

A Delphi study to strengthen research methods training in undergraduate psychology programmes

Robert T. Thibault[1,2†] (0000-0002-6561-3962)

Deborah Bailey-Rodriguez[3] (0000-0002-9931-563X)

James E. Bartlett[4] (0000-0002-4191-5245)

Paul Blazey[5] (0000-0002-8149-9514)

Robin J. Green[6] (0000-0002-2615-4409)

Madeleine Pownall[7] (0000-0002-3734-8006)

Marcus R. Munafò[2,8] (0000-0002-4049-993X)

[1]Meta-Research Innovation Center at Stanford (METRICS), Stanford University, [2]School of Psychological Science, University of Bristol, [3]Psychology Department, Faculty of Science & Technology, Middlesex University, UK, [4]School of Psychology and Neuroscience, University of Glasgow, UK, [5]Faculty of Medicine, Dept of Family Practice, University of British Columbia, [6]School of Psychology and Vision Sciences, University of Leicester, [7]School of Psychology, University of Leeds, UK, [8]MRC Integrative Epidemiology Unit at the University of Bristol

[†]Correspondence should be addressed to Robert T. Thibault; E-mail: robert.thibault@stanford.edu

21 July 2023

Abstract

Psychology programmes often emphasise inferential statistical tests over a solid understanding of data and research design. This imbalance may leave graduates under-equipped to effectively interpret research and employ data to answer questions. We conducted a two-round modified-Delphi to identify the research methods skills that the UK psychology community deems essential for undergraduates to learn. Participants included 103 research methods instructors, academics, students, and non-academic psychologists. Of 78 items included in the consensus process, 34 reached consensus. We coupled these results with a qualitative analysis of 707 open-ended text responses to develop nine recommendations for organisations that accredit undergraduate psychology programmes—such as the British Psychological Society (BPS). We recommend that accreditation standards emphasise (1) data skills, (2) research design, (3) descriptive statistics, (4) critical analysis, (5) qualitative methods, and (6) both parameter estimation and significance testing; as well as (7) give precedence to foundational skills, (8) promote transferable skills, and (9) create space in curricula to enable these recommendations. Our data and findings can inform modernised accreditation standards to include clearly-defined, assessable, and widely-encouraged skills that foster a competent graduate body for the contemporary world.

Keywords. Delphi, psychology education, research methods, consensus, British Psychological Society, accreditation standards, undergraduate, replication, statistics, qualitative methods, research design.

1. Introduction

For graduates from psychology programmes to thrive, they must become effective thinkers in a data laden world. Research methods education in psychology programmes, however, often emphasises inferential statistical tests over a deep understanding of data and research design (TARG Meta-Research Group, 2022), which could lead to the problematic use and interpretation of statistics. Moreover, open research practices are not yet embedded in many curricula and qualitative research methods often remain under-emphasised. In the UK, the British Psychological Society (BPS) sets the requirements for the vast majority of undergraduate psychology programmes through their accreditation standards (British Psychological Society, 2019). Here, we conducted a consensus process, open to the UK psychology community, with the aim to strengthen the research methods section of the BPS accreditation standards.

In response to the 'replication crisis', psychology researchers have increasingly adopted open sciences practices and considered statistical power, sample size, and the use of estimation (Cumming & Calin-Jageman, 2016). These advances, however, are not yet well reflected in psychology curricula. At least five studies have assessed the content of university level psychology programmes in the United States and generally conclude that there have been few updates to the curricula over the past two or three decades (Aiken et al., 1990, 2008; Anglin & Edlund, 2020; Friedrich et al., 2000, 2018). A similar study in the UK found that only 19% of universities had publicly available syllabi describing the content taught in each of their statistics modules in undergraduate psychology (TARG Meta-Research Group, 2022). Although these syllabi rarely contained a lesson-by-lesson breakdown, most mentioned specific inferential statistical tests (e.g., ANOVAs), about half mentioned probability and randomness, effect size, and statistical power, and few mentioned concepts such as confidence intervals, multiple comparisons, meta-analysis, replication, Bayesian statistics, frequentist statistics, and practical significance. Another study surveyed psychology students and instructors in the UK and found that few courses teach alternatives to null hypothesis significance testing and that students' anxiety around mathematics and statistics hold them back (Field, 2014). Notably, a British Academy report highlights this issue in stating that *"A co-ordinated and continuous effort at improving quantitative skills across all phases of education and employment, in all four nations of the UK, is therefore now urgently needed."* (The British Academy, 2015).

Qualitative research skills have been part of the British Psychological Society (BPS) accreditation standards since 2004. However, there is limited research on how they are taught, and instructors may need additional training and resources to effectively teach qualitative methods (Gibson & Sullivan, 2018; Hugh-Jones et al., 2012; Wiggins et al., 2016). Some people also view qualitative research methods as the alternative and 'lesser' approach to quantitative approaches, thus affecting how they are taught (Gibson & Sullivan, 2018; Hugh-Jones et al.,

2012). Taken together, the time is ripe to modernise the teaching of quantitative and qualitative research methods in psychology programmes.

We conducted the present consensus process with the goal of informing an updated version of the BPS *Standards for the accreditation of undergraduate, conversion and integrated Masters programmes in psychology*, specifically Section 2.1.4g of this document (British Psychological Society, 2019). Similar consensus processes have been used to develop standards in over 200 medical education programmes (Humphrey-Murto et al., 2017). The updated accreditation standards could contain specific actionable items for UK psychology research methods curricula, that reflect the need for data skills in the modern world and are adapted to the evolving educational landscape. Our results also provide a foundation for organisations beyond the BPS who seek to modernise research methods education in undergraduate psychology programmes.

1.1 Study objective

To achieve consensus regarding the accreditation standards for research methods education in undergraduate psychology programmes in the UK. Specifically, to provide the BPS with information to update their accreditation standards for research methods education.

2. Methods

We used a modified-Delphi technique to systematically elicit anonymous, asynchronous, and iterative input from a range of stakeholders (Humphrey-Murto et al., 2019). The Delphi study was co-developed with input from the BPS to ensure that the results could be integrated into their *Standards for the accreditation of undergraduate, conversion and integrated Masters programmes in psychology.* The study was approved by the School of Psychological Science Research Ethics Committee at the University of Bristol (ID: 13394). We preregistered a protocol before advertising the study (https://osf.io/5h7bu). Deviations from the preregistered protocol are outlined in Supplementary Material A.

The study included a preparatory stage and two Delphi rounds where participants rated items and provided open-ended feedback (see Figure 1). In a modification to the traditional Delphi method, our Delphi did not include an idea generation round. The preparatory stage was conducted by the Delphi Steering Committee (see next section). They defined consensus, developed a set of questions that would inform updates to the BPS guidelines, and identified

stakeholder groups. The steering committee conducted a non-systematic literature review[1] on research methods education in psychology programmes (see Supplementary Material B). This review served as a basis for formulating the survey questions and was shared with participants at the survey's outset to support informed responses.

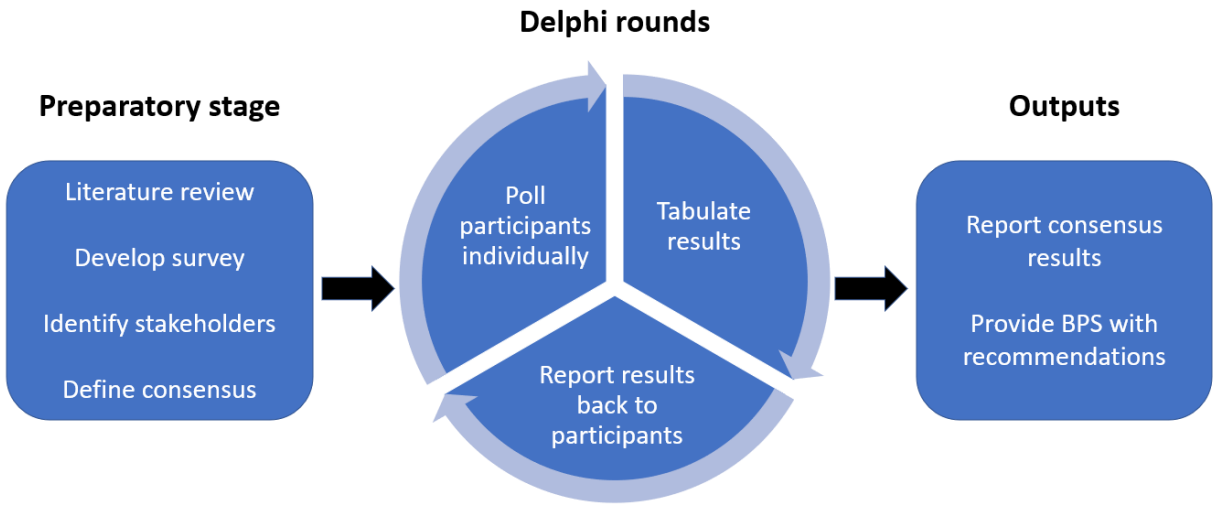


Figure 1. An outline of our modified-Delphi study. A steering committee completed the preparatory stage, participants completed two Delphi rounds, and results were presented to the BPS Undergraduate Education Committee.

2.1 Delphi Steering Committee

We assembled a Steering Committee to guide this Delphi study. We aimed to include members who met a range of representation criteria. These included representation from the following groups: BPS Undergraduate Education Committee, BPS Partnership and Accreditation Committee, Undergraduate Psychology Programme Director, quantitative research methods instructor/expert, qualitative research methods instructor/expert, psychology researcher, statistician, education expert, UK country other than England, non-academic psychologist, and relatively recent graduate from an undergraduate psychology programme in the UK. With these criteria in mind, the Steering Committee comprised 5 members[2] including the project lead (Dr.

[1] Author RTT drafted a literature review based on the TARG Meta-Research Group (2022) article he co-authored. The draft was then shared with all members of the Steering Committee, who suggested several additional papers to include. We also used a snowball method, where we scanned for relevant citations in the publication we reviewed. We did not conduct a systematic review or use a formalised search method.

[2] The Steering Committee originally had 6 members, including Dr. Andy Field–the writer of several statistics textbooks commonly used in undergraduate psychology education. However, due to other commitments Dr. Field's involvement ended before the protocol was finalised.

Robert Thibault), a BPS representative (Dr. Robin Green), a quantitative psychology research methods instructor with experience as a Psychology Undergraduate in the UK (Dr. James Bartlett), an open science and pedagogic expert with experience as a Psychology Undergraduate in the UK (Dr. Madeleine Pownall), and a qualitative psychology research methods expert (Dr. Deborah Bailey-Rodriguez). We unfortunately did not succeed in achieving representation from a non-academic based in psychology. The Steering Committee communicated with the BPS Accreditation Operations Manager (Patricia Lyons) and Chair of the BPS Undergraduate Education Committee (Simon Goodson), to ensure that the study was designed in a way that the results could effectively inform an update to the accreditation standards.

2.2 Participants

We advertised the Delphi survey via mailing lists and social media. We specifically reached out to the BPS, UK Reproducibility Network, ReproducibiliTea, and The Framework for Open and Reproducible Research Training (FORRT) to help advertise. Supplementary Material C contains a template of the invitation text. We did not impose a limit on the number of participants.

Participants needed to meet two criteria. First, they needed to be a member of at least one of the following stakeholder groups: (a) Student in psychology (undergraduate student, graduate student, or non-student who completed their undergraduate degree less than 3 years ago), (b) Research methods instructor in psychology, (c) Academic based in psychology, or (d) Non-academic working in psychology. Participants were instructed to select option (a) if they were an undergraduate or graduate student, regardless of whether they taught or did research; and to select option (b) if they teach or coordinate psychology research methods in an undergraduate or Masters conversion programme, regardless of whether they are also an academic or non-academic psychologist (verbatim instructions are provided in Supplementary Material B). Second, participants needed to either be: (a) based in the UK (or have been a member of one of the four stakeholder groups in the UK within the past 3 years), or (b) be based outside of the UK but be associated with a BPS accredited psychology programme (e.g., the BPS accredits some psychology programmes outside the UK).

Before presenting the Delphi items, the survey presented registration questions. These asked the participants whether they primarily conduct qualitative or quantitative research, are a research methods expert, an undergraduate programme director, and live in the UK; as well as the sector in which they are employed (Supplementary Material D). No incentives were offered for participation.

2.3 Definition of consensus

Before launching the survey, the Steering Committee defined consensus as at least 75% of participants in each non-student stakeholder group rating an item as 'essential' (i.e., between 7-9 on the 9-point scale). A 75% threshold is commonly used in Delphi studies (Diamond et al., 2014) and we felt this was a balanced approach. We excluded students from the definition of consensus because they may not have been exposed to many of the concepts presented in the Delphi. The Steering Committee further made the *a priori* decision that if a stakeholder group had fewer than 12 participants, we would not require that group to achieve 75% essential responses—but instead, we would require that all non-student ratings collapsed together reached 75%. We made this decision to avoid a situation where a very small number of participants are responsible for consensus not being reached.[3] We performed a sensitivity analysis that required 75% essential ratings from all four stakeholder groups (including students), and it revealed no difference in the items that reached consensus.

2.4 Survey

To conduct the survey, we used the DelphiManager software provided by the COMET Initiative.

In Round 1 of the Delphi, participants were asked to rate their level of agreement to 72 items on a scale of 1-9, where 1-3 is 'not important' 4-6 is 'important, but not essential', and 7-9 is 'essential' (see Supplementary Figure 1 for a screenshot of Round 1). These items ranged from specific content to teach (e.g., effect sizes, reflexive practice) to ways of teaching (e.g., evaluation methods) and encouraged resources (e.g., freely available software)—see Supplementary Material F for a complete list of items.

Round 1 comprised 8 blocks which each presented between 5 to 15 items covering the following domains: (1) statistical analyses, (2) quantitative data skills, (3) quantitative research methods concepts, (4) qualitative research methods, (5) research design, (6) reproducibility and open science, (7) accessibility of resources, and (8) miscellaneous. The blocks were presented in a random order. The items within each block were always presented in the same order (as required by the DelphiManager software). The research design block included an attention check item which asked participants to select '3'.

The motivation for this study came from the lead author (RTT) and senior author's (MRM) reflections about the quantitative abilities of psychology graduates, as well as shortcomings in the reproducibility of quantitative psychology research. Items were selected based on previous studies on research methods education in psychology (e.g., TARG Meta-Research Group, 2022)

[3] For example, to prevent a situation where 3 of 10 participants in one stakeholder group rate an item as non-essential (which would preclude consensus)—but the vast majority of a hundred participants across all other stakeholder groups rate an item as essential.

and with the aim of addressing the knowledge and skills gap that leads to irreproducible research. We aimed to word the items in such a way that the BPS could easily integrate the Delphi results into an updated version of the BPS accreditation standards. Moreover, we aimed to make the items specific enough that someone could assess whether that standard is being met. For example, instead of asking if students should learn to 'critically evaluate research', we asked whether students should learn how to 'define and explain questionable research practices (QRPs)' or 'cognitive biases (e.g., confirmation bias)'. Qualitative research methods skills were subsequently included in this Delphi upon a recommendation from the BPS representatives. The lead author (RTT) drafted an initial list of items which the Steering Committee and BPS representatives modified and added to.

After participants rated all items in Round 1, the survey asked them to propose additional items that they felt the survey did not include, but that they would deem important. The Steering Committee added some of these suggested items to Round 2 (many were reworded or combined to better match the Delphi structure). To ensure Round 2 did not take too long to complete, only a few suggested items were added. Suggested items were not added if they did not apply to psychology broadly, overlapped substantially with items in Round 1, or were too vague to be meaningfully integrated into the accreditation standards.

All participants who began Round 1 were invited via email to participate in Round 2. Questions that reached consensus in Round 1 were removed from Round 2. In Round 2, participants were provided with feedback about other participants' responses from Round 1 and asked to re-rate each item. For each item, they were shown the distribution of ratings from each stakeholder group, alongside their own response from Round 1 (see Supplementary Figure 2 for a screenshot of Round 2).

2.5 Open-ended questions

Participants were invited to provide open-ended written feedback at several points during the Delphi study. In chronological order these were: (1) when rating items in Round 1, participants had the option to provide written feedback on each item. (2) After rating all items in Round 1, participants were asked to suggest additional items for Round 2. (3) After completing Round 1, participants were prompted to provide any thoughts they have about the Delphi study. (4) After rating all items in Round 2, participants were asked to give a reason for all their answers that changed rating categories between Round 1 and Round 2. (5) After completing Round 2, participants were prompted to provide any thoughts they have about the Delphi study.

2.6 Analyses

For each item in the Delphi, our main results present the percentage of all participants who rated an item as *essential,* the mean rating across all participants, and whether the item reached consensus. Our open data (https://osf.io/hpsq4/) includes 6 summary statistics datasheets: one for each stakeholder group individually, one for all non-student stakeholder groups collapsed together, and one for all four stakeholder groups collapsed together.

To identify common patterns, one team member (DBR) applied an inductive thematic analysis (Braun & Clarke, 2006), based on a critical realist ontological stance, to the open-ended textual responses. The thematic analysis included a careful reading and re-reading of the textual data followed by thorough line-by-line coding. Emerging themes were then identified and reviewed, and were subsequently grouped together and given labels. A thematic map was produced, which was further refined, and finally the analysis was produced (Braun & Clarke, 2006).

Upon viewing the data, we became aware that the DelphiManager software outputs data in a format that does not allow the registration information (as described in the *Participants* section of this methods section) to be linked to the ratings of the survey items. Thus, we cannot analyse the ratings in relation to the participants registration information. This data structure also leaves us with registration information for only two samples: (1) all participants who at least completed the registration, even if they did not rate any items or failed the attention check, and (2) all participants who at least began Round 2.

3. Results

3.1 Participants

Figure 2 provides a flowchart of participant inclusion. 170 participants began Round 1. Items reaching consensus after Round 1 were based on the 139 participants who passed the attention check and completed Round 1. Our final sample comprised 103 participants who passed the attention check and completed both Round 1 and Round 2. The open data includes datasheets for each of: the final sample, the participants who completed at least one round, and all participants who at least began Round 1.

Table 1 outlines participant characteristics[4]. The majority of participants who began Round 2 primarily do quantitative research and are employed by universities. They registered for the

[4] We preregistered that we would remove participants who responded that they were not associated with the UK. Four participants from our initial sample and one from our final sample selected this response. However, because the participant registration information could not be linked with the participant ratings (as explained in the *Analyses* section of the methods), we could not exclude these participants' ratings.

Delphi using 64 unique email address domain names which included 55 distinct UK academic domain names (i.e., ending in ".ac.uk"). For all participants who at least registered for the study, these numbers were 86 and 69, respectively.

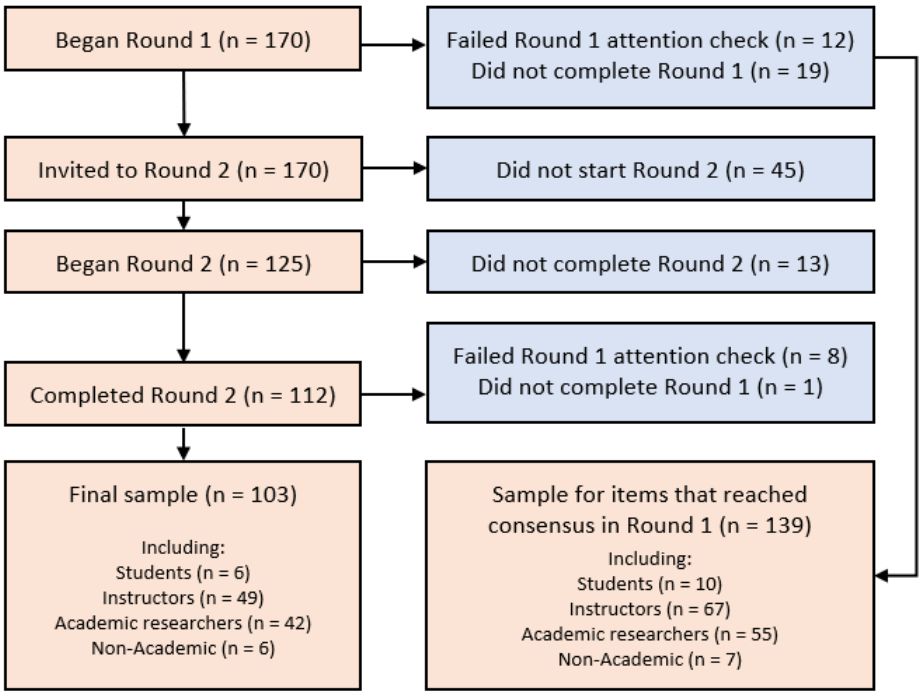


Figure 2. Flowchart of participant inclusion in the final sample. The feedback presented during Round 2 came from the participants in the bottom right box 'Sample for items that reached consensus in Round 1'. Blue boxes indicate participants that were excluded from the final sample.

Table 1. Characteristics of participants who at least began Round 2

Participant charactristic	n = 125
Expertise	
Primarily do quantitative psychology research	77
Primarily do qualitative psychology research	25
Do similar level of quantitative and qualitative research	20
Self-identified as research methods expert	40
Undergrad Programme Director	20
Primary employer	
University (equal teaching and research)	57
University (primarily teaching)	36
University (primarily research)	13
Graduate student	7
Clinical practice	4
Other	4
Undergraduate student	3
Industry	1
Relation to UK Psychology	
Live in UK	123
Associated with BPS accredited programme outside the UK	1
Not associated with the UK	1

Participants could select multiple responses for the 'Expertise' characteristic.

3.2 Delphi item ratings

26 items reached consensus in Round 1 and were not included in Round 2. The Steering Committee added 7 items to Round 2, which thus contained a total of 52 items[5] .

8 items reached consensus in Round 2. 44 items did not reach consensus. Two of the items reaching consensus in Round 2 were added after Round 1 was complete, and 2 others had already reached consensus among the final sample of 103 participants (but not among the 139 who completed Round 1). Thus, among the final sample, only 4 items shifted from not reaching consensus in Round 1 to reaching consensus in Round 2. Across all 103 participants, regardless of stakeholder group, the median percent of participants that rated an item as essential was 74% (interquartile range: 60% to 86%). The median of the average (mean) rating across the 78 items was 7.4 (IQR: 6.8 to 7.9).

Table 2 presents the results for each of the 78 items. To explore these results, we recommend opening the spreadsheet available at https://osf.io/57mbd. This spreadsheet includes the

[5] 72 (Round 1 items) - 26 (items that reached consensus) - 1 (attention check) + 7 (new items) = 52 (Round 2 items).

verbatim items and domains, which could not easily fit in a pdf, but which contain important keywords. For example, whereas some items asked if students should learn how to *apply* a technique, others asked if students should learn to *define* a concept.

Table 2. Rating and consensus for the 78 Delphi items

Item	Domain	Rated essential (percent)	Mean rating	Consensus
Formulate a research question	design	99	8.8	1
Identify and assess ethical issues	design	99	8.6	1
Descriptive statistics	stats	98	8.7	1
Design a study	design	97	8.5	1
Time and support to improve research instructor skills	misc	96	8.5	1
Significance tests	stats	96	8.4	1
Represent data visually	data	95	8.4	1
How descriptive statistics differ from inferential statistics	quant	95	8.3	1
Identify and categorise different types of data	data	94	8.4	1
Sources of bias	OS	91	8.0	1
Research misconduct	OS	90	8.0	1
Generalisability and robustness	OS	89	7.8	1
Practical significance	quant	89	8.1	1
Clean data	data	88	8.0	1
Use descriptive statistics before inferential statistics	data	87	8.0	1
Create a sampling plan and data collection plan	design	86	7.9	1
Effect sizes	stats	86	8.0	1
Methods to assess statistical assumptions	stats	85	7.9	1
Questionable Research Practices (QRPs)	OS	85	7.7	1
Regression	stats	85	8.0	1
Critically appraise qualitative research	qual	83	7.7	1
Explain research question vs hypothesis	design	83	8.2	1
Follow accepted reporting guidelines	design	82	7.8	1
Replication studies and reproducibility	OS	82	7.7	1
Probability and randomness	quant	81	7.7	1
Parameter estimation (95Identify basic study designs	design	95	8.2	2
Assess validity and reliability	design	91	7.8	2
Anonymize data	misc	86	7.9	2
Option for qual, quant, or mixed-methods final year project	misc	85	8.0	2
How exploratory research and confirmatory research differ	quant	85	7.8	2
Operationalize all elements of a study	design	83	7.7	2
Apply experimental and non-experimental research designs	design	82	7.7	2
Cognitive biases	OS	80	7.7	2
Search and collate published research	misc	77	7.4	no
Select a sample size for qualitative research	qual	76	7.4	no
Employ methods known to reduce 'statistics anxiety'	misc	75	7.6	no
The 'replication crisis'	OS	75	7.5	no
Demonstrate understanding of qual data analysis methods	qual	74	7.3	no
Perform qualitative analysis	qual	73	7.3	no
Demonstrate general computer skills for research	data	73	7.3	no
Data, code, and material sharing	OS	72	7.1	no
The existence of different statistical approaches	quant	72	7.2	no
Use reflexive practice	qual	72	7.5	no
Demonstrate understanding of mixed methods research	qual	71	7.1	no
Demonstrate understanding of qualitative frameworks	qual	71	7.2	no
Higher staff to student ratiofor research methods modules	resources	70	7.2	no
Provide syllabi with week-by-week module outline	resources	70	7.2	no
Systematic reviews and meta-analysis	OS	70	7.0	no

Consider diverse perspectives when designing a study	design	69	7.1	no
Emphasize skills that transfer beyond academic research	resources	67	7.1	no
Never entirely grad using closed-book exams	resources	67	7.1	no
Use GUI statistical analysis package	data	66	6.7	no
Collect qualitative data	qual	65	7.2	no
Design a survey	design	65	7.0	no
Psychometrics	quant	65	6.8	no
Apply blinding and randomization	design	63	6.9	no
Perform sample size calculations	design	62	7.0	no
Philosophy of science	OS	59	7.0	no
Preregistration and Registered Reports	OS	59	6.8	no
Allowed to perform a replication as their final year project	misc	56	6.6	no
Explain philosophical underpinnings of qual research	qual	51	6.7	no
Apply qualitative frameworks	qual	50	6.4	no
The publication process	OS	50	6.3	no
Meta-research / meta-science	OS	47	6.4	no
Allowed to conduct their final year project in a team.	misc	46	5.8	no
Determine a smallest effect size of interest (SESOI)	design	44	6.5	no
Reward structures in research and academia	OS	44	6.5	no
The existence of different statistical frameworks	quant	40	5.9	no
Equivalence testing	stats	34	6.0	no
Factor analysis	stats	33	5.8	no
Only use freely available software	resources	28	5.4	no
Use a programming language	data	23	5.3	no
Preregister quantitative aspects of final year project.	misc	22	5.1	no
Multiverse analyses / many-analyst approaches	OS	17	4.8	no
Alternative measures of effect sizes	quant	16	4.9	no
Make syllabi publicly available	resources	14	4.5	no
Simulate data	data	9	4.4	no

Items are ordered by the column 'Consensus' then 'Rated essential (percentage)'. Many items have been paraphrased so they can fit in this table. The domains have been shortened (stats = statistical analyses; data = quantitative data skills; quant = quantitative research methods concepts; qual = qualitative research methods; design = research design; OS = reproducibility and open science; resources = accessibility of resources; misc = miscellaneous). Full verbatim descriptions of the items and domains are available in the following spreadsheet: https://osf.io/57mbd. The rating scale ranges from 1 to 9, where 1-3 is "not important" 4-6 is "important, but not essential", and 7-9 is "essential". The consensus column contains a value of 1 if consensus was reached in Round 1, 2 if reached in Round 2, and 'no' if consensus was not reached. Items that reached consensus in Round 1 have the columns 'Rated essential (percent)' and 'mean rating' taken from Round 1 (because these questions were not included in Round 2).

We performed three sets of sensitivity analyses. We assessed differences in ratings between (i) Round 1 and Round 2, (ii) the instructor and academic stakeholder groups, and (iii) the initial sample of 170 participants and the final sample of 103 participants (see Table 3). We did not compare results from the students and non-academic stakeholder groups because they had few enough participants that the comparisons would be uninformative or potentially misrepresentative. Relatively small differences in ratings occurred between rounds and between the instructors and academic stakeholder groups. The open-ended responses revealed several reasons for why participants changed their ratings between rounds, including viewing the ratings of other participants, reflecting further, discussing with colleagues or students, and gaining further knowledge after Round 1. Results were very similar between the

sample of participants that at least began Round 1 (170 participants) and the final sample (103 participants).

Table 3. Sensitivity analyses between Delphi rounds, stakeholder groups, and initial and final samples

Differences	Rounds	Stakeholders	Sample
Reached consensus (n item)	4	6	1
Rated essential (percentage)	3.8	5.3	1.3
Rating (mean)	0.15	0.24	0.06

Of the 47 items included in both Round 1 and Round 2: Four went from not reaching consensus to reaching consensus, the median absolute change in the percentage of participants that rated an item as essential was 3.8%, and the median absolute difference in participants' rating of an item was 0.15. We found slightly larger differences between the final ratings from the instructor and academic stakeholders groups. Only small differences existed between ratings from the sample of participants who at least began Round 1 and the final sample.

3.3 Consensus summary results

In this section, we block related items into 13 overarching findings (see Table 4). We conceived these blocks based on whether the items reached consensus and whether they can be interpreted together to help formulate a specific recommendation[6]. In brackets, we present the percentage of all participants who rated the items as *essential*.

Table 4. Recommendation topics derived from consensus results

Consensus largely reached	Consensus reached for some items	Consensus largely not reached
Understanding data	Qualitative methods	Advanced analysis techniques
Research design (general)	Reproducibility and open science	Research design (specific)
Descriptive statistics		Approaches to research
Inferential statistics		Computer skills
Critical assessment		Module format
		Final year projects

[6] Some of these blocks overlap with how the survey was presented: in 8 blocks, or 'domains'. However, whereas those domains were based only on the overarching topic of the items, the blocks in this section incorporate both the topic and the consensus results.

3.3.1 Consensus largely reached

A. **Understanding data.** Consensus was reached that students should learn how to identify and categorise different types of data (94%), clean data (88%), anonymise data (86%), and represent quantitative data visually (95%).

B. **Research design (general).** Consensus was reached that students should learn how to formulate a research question (99%), design a study to answer a specific research question (97%), explain the difference between a research question and a hypothesis (83%), and identify basic study designs (95%). Consensus was also reached for learning how to create a sampling plan (86%), operationalise all elements of a study (83%), and apply experimental and non-experimental research designs (82%).

C. **Descriptive statistics.** Consensus was reached that students should learn how to calculate descriptive statistics (98%), explain the importance of descriptive statistics and how they differ from inferential statistics (95%), and use descriptive statistics effectively before learning to perform inferential statistical tests (87%).

D. **Inferential statistics.** Consensus was reached that students should learn how to calculate significance tests (96%), regressions (85%), effect sizes (86%), and parameter estimations (e.g., confidence intervals) (79%). Consensus was also reached that students should learn to define and explain probability and randomness (81%), and was almost reached that students should learn to explain the existence of different statistical approaches, including parameter estimation and significance testing (72%)[7].

E. **Critical assessment.** Consensus was reached that students should learn how to identify and assess ethical issues (99%), assess validity and reliability (91%), define and explain the difference between statistical significance and practical significance (89%), and define and explain the value of exploratory research and how it differs from confirmatory research (85%). Defining and explaining systematic reviews and meta-analysis (70%) did not reach consensus.

3.3.2 Consensus reached for some items

[7] In the manuscript section 'Consensus summary results' we use the term 'almost reached consensus' for all items with >70% essential ratings but which did not meet our preregistered definition of consensus. This decision was not preregistered. We feel it helps avoid a strict dichotomization.

F. **Qualitative research.** Learning to critically appraise qualitative research reached consensus (83%). Several other qualitative items almost reached consensus, including learning to demonstrate understanding of several methods of qualitative data analysis (74%), perform qualitative analysis (73%), use reflexive practice (72%), demonstrate understanding of various qualitative frameworks (71%)[8], and demonstrate understanding of mixed methods research (71%). Some qualitative items did not reach consensus, including learning to collect qualitative data (65%), explain the philosophical underpinnings of qualitative research (51%), and apply qualitative frameworks in their own research (50%). 15 participants provided open-ended feedback to specific Delphi items on qualitative methods, and 19 participants provided general comments that mentioned qualitative methods. These textual data are analysed in the manuscript section *Theme 1: Important factors.*

G. **Reproducibility and open science.** Several items related to reproducibility and open science reached consensus, including that students should learn to define and explain sources of bias (91%), cognitive biases (80%), questionable research practices (QRPs) (85%), generalisability and robustness (89%), research misconduct (90%), and replication studies and reproducibility (82%).

   Other items that almost reached consensus include learning to define and explain the replication crisis (75%)[9], and define and explain data, code and material sharing (72%). Items more focused on the process of research did not reach consensus, including learning to define and explain the publication process (50%), reward structures in research and academia (44%), preregistration and Registered Reports (59%), and meta-research (47%).

3.3.3 Consensus largely not reached

H. **Advanced analysis techniques.** More advanced analysis techniques did not reach consensus, including learning to perform equivalence testing (34%), perform factor analysis (33%), simulate data (9%), and explain multiverse analyses / many-analyst approaches (17%).

I. **Research design (specific).** Items on research designs specific to certain study types did

[8] Including the term 'various' in this item may have prevented some participants from rating it as essential.

[9] Some items had >75% essential ratings when collapsed across all participants, but did not have >75% agreement within each stakeholder group, and thus did not meet our definition of consensus.

not reach consensus, including learning how to design a survey (65%), apply blinding and randomisation (63%), and explain psychometrics (65%). Items about sample size and effect sizes also did not reach consensus, including learning to perform a sample size calculation for quantitative research (62%), determine a smallest effect size of interest (44%), and explain alternative measures of effect sizes (e.g.; probability of superiority) (16%). Learning to select a sample size relevant to the qualitative method being used (76%) almost reached consensus.

J. **Approaches to research.** Consensus was not reached that students should learn how to consider diverse perspectives when designing a study (69%), define and explain philosophy of science (59%), or define and explain the existence of different statistical frameworks; including frequentist statistics and Bayesian statistics (40%).

K. **Computer skills.** Consensus was not reached regarding whether students should learn how to use a programming language to manage and analyse data (23%) or use a statistical analysis package with a graphical user interface (66%). Whether research methods modules should only use freely available software (28%) did not reach consensus. Consensus was almost reached for learning to demonstrate general computer skills for research (73%) and search and collate published research (77%).

L. **Module format.** No items regarding the format of modules reached consensus, including to never entirely grade with closed-book exams (67%), emphasise skills that transfer beyond an academic research context (67%), have a higher staff to student ratio than for non-research methods modules (70%), provide students with syllabi that include a week-by-week outline of the module contents (70%), or make syllabi publicly available (14%). Actively employing teaching and grading methods known to reduce 'statistics anxiety' (75%) almost reached consensus.

M. **Final year projects.** Consensus was reached that students should have the option to conduct a qualitative; quantitative; or mixed-methods project in their final year research (85%). Consensus was not reached on whether students should preregister the quantitative aspects of their final year project (22%), be allowed to perform a replication as their final year research project (56%), or be allowed to conduct their final year project in a team (46%).

3.4 Thematic analysis of the open-ended questions

Participants provided written responses to open-ended questions. These included feedback from 41 participants on 222 specific items; from 30 participants for 82 suggested items; and from 73 participants regarding 337 ratings that crossed a rating boundary (e.g., from 'important, but not essential' to 'essential'). A small number of participants were responsible for large portions of these open-ended responses (shown in Supplementary Table 1). 45 participants left a general comment after Round 1, and 21 participants left a general comment after Round 2.

The thematic analysis addressed the research question: what are important issues to consider when designing the research methods curriculum? The analysis generated two main themes: 'important factors' and 'constraining factors', which encompass the key issues in the design of research methods curricula. Figure 3 illustrates these themes and corresponding sub-themes.

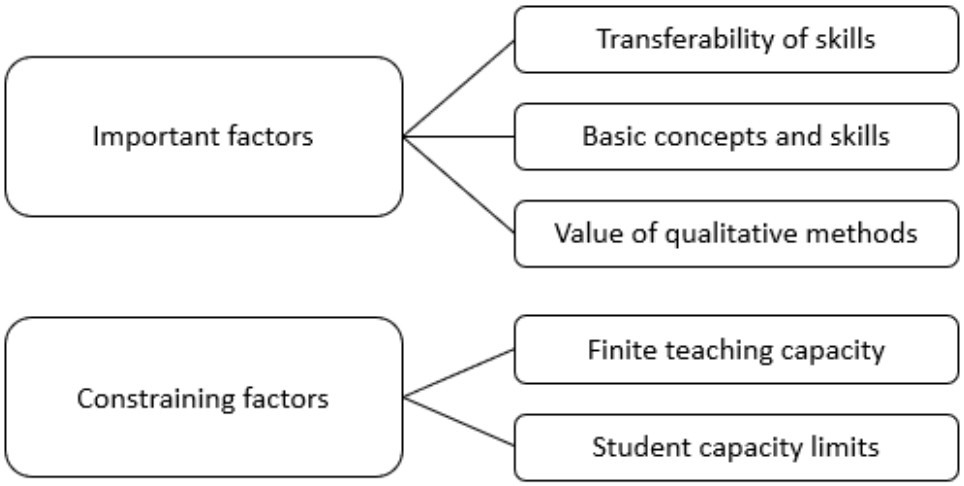


Figure 3. Thematic map with two main themes, and their sub-themes.

3.4.1 Theme 1: Important factors

This theme delves into the various factors that were considered to be important to include in the research methods curriculum design. The transferability of skills to the workplace, ensuring students understand basic concepts and skills, and the value of qualitative research were prominent issues, and these sub-themes are explored below.

Transferability of skills

We identified the transferability of skills to employment contexts as an important factor to be accounted for in the research methods curriculum design. For example, a participant said: *"A balance is needed to future-proof student skills while focusing on practical understanding of*

research methods (at UG [undergraduate] *level at least).”* Another participant wrote: *“there should be more consideration given to the* […] *skills that might transfer to other fields, outside academia.”*

Most graduates do not pursue an academic career after their psychology degree (Palmer et al., 2021). Thus, teaching research methods skills that apply beyond academia would be important. This issue was highlighted by a participant:

> *“It would also be great to see research methods taught in a way that emphasizes real-world applications for the students. The vast majority of psychology undergraduates do not go into Psychological research so, while I appreciate the drive to make University level research methods training more consistent with the current state of psychological research, I am concerned that we might be missing the type of data handling skills that students would use in careers that are data focussed but outside of psychological research.”*

This issue was further exemplified by other participants, for example: *“Most psychology undergraduates are not going to be researchers or undertake research in the future”* and *“It is important for me, as a practicing researcher, to know how to calculate effect sizes, perform a power analysis, navigate publication etc. But is it important for me that students can do these things, given that most of them will not need to in the future?”* Participants understood the significance of teaching research methods in ways that are applicable to non-academic careers.

To this end, participants provided views about the inclusion of transferability skills when teaching research methods that could be applied to non-academic employment settings, for example: *“Teach what is used for research and industry”* and *“applicability to non-academic settings is important”.* With regards to the questions asked in the survey, one participant wrote:

> *“I would have liked to see some questions about what research methods skills are important for students who graduate and don't get employment in psychology, which is what most of our students do. I feel some of the issue*[s] *are not relevant to someone who is carrying out research in a non-academic context.”*

These responses illustrate the importance and value of designing a research methods curriculum which actively includes research methods training that can be leveraged in non-academic employment routes. Nevertheless, the value of teaching research skills which applies both to academic and non-academic pathways, was another important factor for the curriculum. For example, a participant wrote: *“Asking students to work as a team is more*

realistic to how research is conducted in the real world (whether that be academia or in the public or private sector)".

Furthermore, for psychology graduates intending to follow academic careers, our analysis suggests that it is also important to teach transferable skills that reflect what happens in real-world research, for example: *"we falsely give students the idea research is a fairly solo/small homogenous (student) group activity whereas the truth in many fields related to psychology is that it needs multidisciplinary teams including methodologists and lay advisors".* For further illustration, in relation to the survey question 'the publication process', another participant wrote: *"very important when it comes to following an academic path"*. Additionally, in relation to the survey question 'students should be allowed to perform a replication as their final year research project', another participant stated: *"This is more representative of actual psychology research, and in line with open science practices."* The comments demonstrate the importance of transferable skills taught in line with academic employment routes.

However, there was some tension in which specific skills should be taught as important for transferability, for example a participant stated: *"SPSS is still very important in government, business, and charities",* and another wrote: *"SPSS is in declining use. Psychology departments do students a disservice by continuing to use software that is not used in other disciplines or in industry"*. Despite the disparity in the specific transferable skills some participants suggested, it remains important that skills learnt in the research methods curriculum have applicability to employment contexts, in both academic and non-academic pathways.

Basic concepts and skills

We identified that ensuring students understand basic concepts and skills was another factor deemed to be essential in the research methods curriculum design. For example, participants stated: *"I believe students should learn the basics of all research methods"* and *"we need to focus on students getting the foundational methods knowledge."*

There were concerns about the increasing number of topics and skills being taught in the research methods curriculum at this level, to the detriment of students' understanding and ability to apply these properly. To illustrate, a participant wrote: *"There is quite a focus on teaching UG* [undergraduate] *students more and more topics, and using potentially more technical platforms. I would also like to see emphasis on improving students' understanding and ability to apply foundation concepts",* and another stated *"at the UG level what we need to be doing is covering LESS but with greater rigour and confirmation that students can actually USE and UNDERSTAND what we have taught them."* The responses demonstrate the importance of making sure that the research methods curriculum includes space and time for

the teaching of concepts and skills considered to be foundational, and to avoid the overload of an ever-increasing number of topics and skills when basic understanding of concepts and skills seems to be lacking.

The negative impact of the increased taught concepts and skills is illustrated in the following participants' statements: *"I end up essentially doing the major parts of their* [final year] *project for them"* and *"We teach research methods in 1st year and 2nd year and they arrive at their 3rd year dissertations knowing practically nothing, despite often doing well on earlier exams."* Not ensuring that students comprehend basic research methods concepts and skills in the earlier years appears to have a negative knock-on effect on their ability to understand and conduct their final year projects.

Furthermore, the importance of ensuring a basic understanding of concepts and skills also applies when teaching students to use analysis software. For example, participants wrote: *"For all of these* [software skills] *it's necessary to explain why we do these tests, what they mean, give real world examples"* and *"For too long have psychology departments "taught SPSS" instead of teaching statistical analysis".* It is important that students have foundational understandings so that, when the time comes to use analysis software, they comprehend what they are doing and why. This issue is further exemplified by the following quote: *"Should we address the problematic tendency to emphasise which stats-package dialogue-boxes to tick at the expense of teaching a deep understanding of what the analysis actually does?"* This comment illustrates the importance of designing the research methods curriculum in a way that ensures students have a solid understanding of basic research concepts and skills.

Additionally, our analysis found that ensuring students develop the ability to think critically was another important foundational concept for developing their understanding and appraisal skills. For example, in this context a participant stated: *"I would like the field to be careful about ensuring that students are not simply directed towards a checklist approach",* and another wrote: *"we ought to be teaching ways to understand, critique, appraise and undertake research".*

The value of ensuring students learn to evaluate research transcends the particular topic under critical review, as exemplified by some participants: *"not everyone will want to use qualitative research methods; but everyone should be able to evaluate published qualitative research"*, *"I think they need to be made aware of the limitations of quantitative research, which they tend to think of as sacrosanct"*, *"It's essential for students to understand that this* [the publication process] *is not a neutral process and factors in the process create a biased literature"* and *"I'd prefer them to know why replication is relevant and why it's become a fixation as opposed to other issues."* These responses demonstrate the importance of ensuring

students develop critical thinking skills as foundational in all research-related contexts, and integrated into the curriculum as such.

On the whole, the findings illustrate the importance of designing a research methods curriculum that ensures students understand basic research concepts and skills, including in teaching analysis software and in the development of critical thinking skills.

Value of qualitative methods

Our analysis identified the value of qualitative research and its inclusion in the research methods curriculum as another important factor. Illustrating this finding, participants wrote: *"students need more qualitative methods input, this is crucial to ensuring the quality of qualitative research and of future training"*, *"qualitative analysis is profoundly important"*, and *"I have rated most of these* [survey items relating to qualitative research] *as essential".*

Additionally, it was recognised that the survey further problematised the perception that qualitative methods are not as valuable as quantitative methods, for example: *"It is interesting that even in this study, a greater emphasis is placed on quantitative over qualitative skills (judged by the number of questions). I believe these should be given equal weight."*

Furthermore, our analysis suggests that qualitative research is currently viewed as having less importance than quantitative methods, but several participants wrote that it should have the same value status in the research methods curriculum. For example, participants stated: *"QUANTitative and QUALitative research methods should be given equal attention by staff and students"*, *"There is an imbalance here on the quant to qual methods - both are equally important and should be given equal importance in undergraduate study"*, and *"Qualitative research methodology should be given the same emphasis as quant so that students appreciate the relative value and appropriate applications."* These comments indicate that participants felt that qualitative methods should hold equal value to quantitative research in the curriculum, and that this value should be held by research methods instructors as well as students.

Another important factor that stems from recognising the value of qualitative research, is that some qualitative concepts are directly applicable to quantitative methods. This was particularly salient when participants were responding to the survey question 'Explain the philosophical underpinnings of qualitative research', where some stated: *"And of quantitative research!"*, *"Students need to know about the theoretical and epistemological underpinnings of research full stop, not just qual"*, and *"Should do this in all research"*.

Overall, qualitative methods are considered to be valuable, and an important factor to include in the research methods curriculum design, alongside the view that they should hold equal status to quantitative research methods.

3.4.2 Theme 2: Constraining factors

This theme unpacks several factors that were considered to be constraining in the design of research methods curricula. Finite teaching capacity and student capacity limits were significant issues. These sub-themes are examined below.

Finite teaching capacity

Our analysis identified that the finite amount of time to teach the many research concepts and skills was considered to be a constraining factor. For example, in response to the survey items, participants stated: *"The study doesn't really take into consideration what is achievable in a given timeframe and whilst a lot of ideas and concepts are really important I am not sure how you would fit them all in"*, *"a lot of my observations of things being 'important but not essential' are really shaped by my understanding of the time constraints in our teaching"*, and:

> *"I feel like it's a bit too easy to say 'everything is really important!' (certainly the RM sample looked like almost everyone was saying '9' for loads of the items), when actually there's a finite amount we can teach our students and expect them to learn in Y1 and Y2 of an UG degree."*

Despite participants wanting to support the value of many of the survey items proposed, it seems responses also took into account the finite amount of capacity available to teach these concepts, even though this was not specifically asked as part of the survey. This finding reflects the importance of taking this constraining factor into account when designing the research methods curriculum.

Concerns regarding finite teaching capacity also extended to the teaching of particular research skills such as mixed methods, for example: *"Mixed methods is often not feasible within the timeframe allowed; it may also require careful staff allocation which could create lack of equality of supervision",* and to learning to use a programming language, for example: *"This entails considerable staff time."* These responses illustrate the need for careful consideration as to what concepts and skills should be included in the research methods curriculum, considering the finite amount of teaching capacity.

The impact of what can be taught in the given amount of time available for research methods can be a negative one, both on students and on instructors, as exemplified by the following

participants: *"In my experience, most students have absolutely no idea of what they're doing and have no capacity or time to be trained, either on their part or mine because I am so wildly overwhelmed",* and in response to the survey item 'Have a higher staff to student ratio than for non-research methods modules': *"Definitely would be good, but hard to implement and also ties closely together with how many taught hours are involved (labs/lectures) and what that means for teaching loads."*

On the whole, the finite capacity for teaching research methods is a constraining factor on the concepts and skills that could be included in a research methods curriculum; especially in view of the negative impact overinclusion can have on students and instructors.

Student capacity limits

Our analysis also found that restrictions in student capacity was a constraining factor regarding what is feasible to include in a research methods curriculum. This finding is illustrated by a participant's comment: *"I think at the moment we're trying to turn all our UG students into PhD level R users and researchers (too much!)."*

Participants seemed to feel that there are limits to what students can learn within the degree time frame, for example: *"As much as I use and like R, I'm not sure it's reasonable to make all students learn to code"* and *"I don't think it's necessary for students to apply multiple types of research designs; there is only so much time in a degree. I think it's more important to know about it"*.

Furthermore, participants thought that certain research skills were too advanced for inclusion in the research methods curriculum. Mixed methods were one such concept, for example: *"I worry that mixed methods can be too complex/large scale to be achievable", "mixed methods reports are too complicated for final year research",* and *"Mixed methods is less important than good qual and good quant and throws up some real issues for final year projects."*

Participants also considered learning to use a programming language to be too advanced for students: *"many students may be overwhelmed by learning a programming language (which may increase students leaving courses)"*, *"This would put off and disadvantage many students"*, and *"Sadly, we struggle to get our students to get to grips with SPSS; when have tried to teach R, it has been even more difficult"*. These responses indicate that designing a research methods curriculum requires careful consideration regarding the concepts and skills that fit the capacity of students at this level.

Nevertheless, the option for students to specialise in more advanced research concepts and skills was considered to be a potential way to overcome this constraining factor, for example:

"I'd like to see research methods training that starts more broadly and becomes more specific; offering students a choice of where to specialize." In response to the survey item 'Use a programming language to manage and analyse data', some participants stated: *"We can certainly teach why this would help and encourage those who're interested/able to explore it* […]. *This should be an optional skill"*, and *"This is important but not essential for all students. They should be given an opportunity to learn this as part of an elective module"*

Overall, student capacity limits are also a constraining factor that should be considered when designing research methods curricula. There are limits to what students can learn within the degree time frame, and some research concepts and skills were thought to be too advanced. Providing the option for students to learn more advanced concepts and skills could be a possible solution.

4. Discussion

4.1 Summary of results

Consensus was reached for 34 items. These items spanned topics including data skills, general research design, descriptive statistics, inferential statistics, critical assessment of research, and to some extent qualitative methods, reproducibility and open science. Consensus was not reached for 44 items. These items spanned advanced analysis techniques, specific research designs, approaches to research, computer skills, module formats, and final year projects. A qualitative analysis of open-ended responses highlighted the importance of understanding basic concepts, valuing qualitative research methods, and learning transferable skills; while also acknowledging limits on how much material can fit in an undergraduate programme and how much students can absorb in this finite time period. Taken together, these results can provide valuable information for instructors, programme directors, and organisations that develop accreditation standards.

4.2 Relation to the literature

The consensus results partially overlap with the content that appears on publicly available curricula for quantitative research methods in psychology programmes (TARG Meta-Research Group, 2022). For example, items such as descriptive statistics, inferential statistics, and critical evaluation were prevalent across curricula and also rated highly in this Delphi study. On the contrary, items such as effect sizes, confidence intervals, data cleaning, practical significance, and replication were less prevalent in the curricula[10], but were highly rated in this

[10] These topics may be taught nonetheless. However, they do not appear in the curricula, which may indicate that they are not emphasised.

Delphi study. Adding these topics to accreditation standards presents one mechanism to encourage their adoption. Almost all the curricula mention SPSS; however, learning to use a statistical analysis package with a graphical user interface did not reach consensus in our study. Our qualitative analysis further suggests that participants were concerned that students learn how to 'point-and-click' in SPSS, rather than gain an understanding of the analyses for which they are using the software. Comparing curricula to accreditation standards—and to the results of this Delphi study—can help understand whether the educational content of psychology programmes aligns with community expectations.

Several qualitative items had lower ratings than quantitative items, reflecting a previously observed trend. For example, qualitative methods appear to be underrepresented in curricula and perceived as an alternative and 'lesser' approach to quantitative methods in UK psychology programmes (Gibson & Sullivan, 2018; Hugh-Jones et al., 2012). Interviews with psychology instructors also suggest that some programmes would need additional instructor expertise to effectively teach and supervise qualitative research methods (Wiggins et al., 2016). Indeed, almost all participants in our study (96%) agreed that research methods instructors should be given time and support to improve skills they plan to teach, including qualitative methods. Given the importance the BPS places on qualitative methods (e.g., the Qualitative Methods in Psychology Section of the BPS), the prominence of qualitative methods in the accreditation standards could be raised.

Transferability of research methods skills emerged as a theme and appears relevant given the diverse career paths that psychology graduates follow. A recent report analysing data from the Higher Education Statistics Authority (HESA) found that "There is no common career path for psychology graduates, as they go on to work in a broad array of roles and settings" (Palmer et al., 2021). The report further states that only about 6% of graduates become registered professionals in psychology and that many go into roles in the health sector, retail, administration, public relations, marketing, and human resources. In these roles, a solid foundation in qualitative and quantitative skills likely trumps the ability to perform inferential statistical tests. Even in careers where inferential statistics are necessary, such as academic research, foundational data skills are also necessary. For example, a recent paper calling for UK psychology education to emphasise data skills demonstrated that, to analyse a realistic quantitative dataset in psychology, data wrangling accounts for about 80% of the steps and statistical procedure account for only 20% of the steps (McAleer et al., 2022).

4.3 Recommendations

As per our study objective, we provided the BPS Undergraduate Education Committee with nine core recommendations for updating their accreditation standards (see Supplementary

Material G; also summarised in Box 1). We developed these recommendations by considering the combination of the consensus summary results (section 3.3) and the thematic analysis (section 3.4). Although we would recommend that all 34 items that reached consensus be considered for inclusion in a research methods curriculum, we binned items into nine recommendations that integrate the qualitative data and hopefully facilitate the implementation of the recommendations.

Box 1. Recommendations for research methods education in undergraduate psychology programmes.

1. Require a strong understanding of **data and quantitative data skills.**
2. Emphasise general skills in **research design**.
3. Prioritise a solid foundation in **descriptive statistics**.
4. Provide students with a **framework for critical assessment** of research claims.
5. Raise the prominence of **qualitative methods** throughout the accreditation standards.
6. For **inferential statistics**, require that **parameter estimation techniques**, such as confidence intervals and effect sizes, are taught alongside significance testing.
7. Give precedence to teaching **foundational research methods skills** (as outlined in Recommendations 1-5).
8. Promote content that elucidates how research methods skills **transfer beyond academia.**
9. **Enable Recommendations 1-8** by encouraging research methods education throughout the programme, focusing on fewer skills in greater depth, and offering optional modules for more advanced research methods skills.

Ratings were very high and consensus reached for data skills, basic research design, descriptive statistics, and inferential statistics. Almost 90% of participants rated it essential that students learn to use descriptive statistics effectively before learning to perform inferential statistics. Our qualitative analysis also highlighted the need for students to master foundational quantitative and qualitative skills, rather than attempt to perform analyses that they understand poorly. These findings challenge the null hypothesis significance testing (NHST) centric approach taken in many research methods curricula and suggest that the psychology community places importance on ensuring students develop a deeper understanding of the research skills they are using and why they are using them.

Open-ended comments raised the point that students should learn how to answer a research question and focus on fewer technical abilities. With this in mind, research methods education could adopt a problem-solving approach by teaching students how to ask a clear question, design an effective research plan, identify what data is needed to answer their question, and how that data could be collected (e.g., the Problem-Plan-Data-Analysis-Conclusion model, as suggested by Spiegelhalter, 2019).

Critical assessment is a pillar throughout the 2019 BPS accreditation standards and mentioned in almost all curricula assessed in a previous study (TARG Meta-Research Group, 2022), but specific concepts and tools are generally not outlined. Updated accreditation standards could include specific items, such as learning about replication and sources of bias, to create a more structured approach for critically assessing the psychology literature and other forms of information. This topic is linked to integrating the principles of open science into undergraduate education, which others have encouraged (e.g., Pennington, 2023; Pownall et al., 2023).

44 items did not reach consensus. These items spanned topics including module format, final-year projects, computer skills, approaches to research, and advanced analysis techniques. Many of these items received a high-level of agreement, but fell short of consensus. There was **not** consensus against teaching these items.

4.4 Limitations

Our study design entails limitations on the claims we can make and how they can be interpreted. First, Delphi studies assess the opinions of a community. They do not establish what educational content is most effective. For example, some participants may simply provide low ratings for items they are unfamiliar with. In our results, for example, consensus was reached for learning about practical significance. However, related concepts which participants may be less familiar with received low ratings (e.g., alternative measures of effect sizes and smallest effect sizes of interest). We hope to have mitigated this limitation by providing the option 'unable to rate', which was used for 3% of ratings.

Second, the format and content of our Delphi was specifically designed in relation to the BPS accreditation standards. This meant that we selected items and phrased them in such a way that our results could easily be integrated into these standards. In this sense, we did not present items that challenged core components of the standards (e.g., the inclusion of a final

year project[11]) or content that already exists in the standards which we believed participants would be unlikely to disagree with (e.g., 'critical evaluation'). Many Delphi studies include an initial idea generation round where participants are asked to suggest items before they see or rate any item. Some Delphi studies also include an item prioritisation round, where participants rank the items that reached consensus. Due to limited resources and time constraints, we did not include these rounds.

Third, we targeted four stakeholder groups, but only achieved a substantial number of participants in two of these groups. These two groups overlapped substantially, as most research methods instructors are likely also academic psychologists, and their ratings were relatively similar. Very few students and non-academic psychologists participated. Thus, in Round 2, most participants saw the ratings from only a few students and non-academic psychologists (which may have comprised an unrepresentative sample) as well as from one other group with similar responses to their own group. This combination of factors may have contributed to the limited changes to ratings between Round 1 and Round 2[12] and also resulted in a failure to capitalise on this strength of the Delphi technique.

Fourth, our sample was likely biased towards quantitative psychologists and people who are highly interested or opinionated about research methods. A majority of the participants reported primarily using quantitative methods and few reported primarily using qualitative methods. This distribution of participants—which may or may not reflect the distribution of the psychology community in the UK—could hold responsible for the generally higher ratings

[11] Our findings also raise a larger question about the structure of undergraduate psychology programmes in the UK. These are generally 3-year programmes (or 4-year programmes in Scotland) where students are expected to learn quantitative and qualitative skills, and conduct a "substantial piece of research…[that] typically involves the collection of original empirical data from participants" (British Psychological Society, 2019). Meanwhile, the following themes emerged from our analysis: (i) finite teaching capacity and (ii) student capacity limits, coupled with (iii) the need to improve foundational skills, and (iv) a tension between quantitative and qualitative methods. Some other countries (e.g., United States, Canada), generally offer 4-year psychology programmes offering specialisation as either a Bachelor's of Science or a Bachelor's of Arts, while allowing students to select from a range of optional modules and only requiring substantial final-year projects from students in Honour's programmes. The results from our present study brought our attention to these differences, although this study does not (and was not designed to) shed light on the benefits and drawbacks of programme structures in different countries.

[12] We also removed items that reached consensus in Round 1, which made it impossible for an item to go from consensus to non-consensus. We did this to ensure Round 2 didn't take too long to complete (in Round 2 participants also saw a graph for each item). We are not concerned that many items not presented in Round 2 would have lost consensus because most of these items had high percentages of essential ratings, ratings were generally stable across rounds, and all ratings with a percentage of essential ratings above 55% in Round 1 increased this percentage in Round 2 (except one item that had a 1% decrease).

for quantitative items as compared to qualitative items. By design, our study also reflects only the views and priorities of the UK psychology community.

Fifth, the study was originally conceived to ask only about quantitative issues, and was thus weighted toward quantitative methods. Qualitative methods were included upon the suggestion of BPS representatives. We did not include items asking about the proportion or ordering of teaching quantitative versus qualitative methods. Several participants provided feedback expressing concern regarding this quantitative-qualitative imbalance and stated that some qualitative items were poorly worded. This imbalance may have impacted the distribution of ratings between the quantitative and qualitative items.

Sixth, participants may have overlooked conditional words that preceded some items. For example, one question asked if students should learn how to '*calculate/perform* significance tests', and another asked if students should learn to '*define and explain* systematic reviews and meta-analysis'. This oversight could have lowered ratings for some items because participants may have thought that these questions were asking if students needed to learn how to *perform* a meta-analysis, for example.

We were aware of the limitations of the modified-Delphi format we used before beginning the study and deemed them acceptable. The shortcomings of sampling bias and limited student and non-academic engagement limits the generalisability of our results to the psychology community at large. Nonetheless, our data and findings provide a resource that can help inform accreditation bodies, programme directors, and module instructors about what the UK academic community believes is essential for undergraduate psychology students to learn.

Our study also had several strengths. The Delphi method is a recommended method to produce guidelines on topics where data is scarce and expert-opinion is the best available evidence. Our mixed-methods approach also provides a robust understanding of participants' opinions. We worked with representatives from the BPS to ensure that we selected items and worded them in a way that facilitates integration into updated accreditation standards. We also received responses from over 100 members of the UK psychology community from more than 50 UK universities, and including 20 programme directors. Finally, all the raw data, summary datasheets, and analysis code are publicly available for others to explore.

4.5 Conclusion

Our study provides datasets, both quantitative and qualitative, on the research methods skills that UK-based instructors and academic psychologists deem essential for undergraduate psychology students to learn. Our findings suggest widespread agreement that research methods education in undergraduate psychology should emphasise foundational skills in

research design, data handling, statistics, qualitative methods, and critical assessment, while providing students with transferable skills and not overloading them with advanced techniques. Organisations that create educational standards for psychology programmes—such as the BPS—can draw on our findings to help develop broadly-accepted and clear-cut expectations for research methods education. Such initiatives could foster cohorts of graduates with an established set of competencies tuned for the contemporary world.

Ethics approval

Ethical approval was obtained before advertising the study and launching Round 1 from the School of Psychological Science Research Ethics Committee at the University of Bristol (approval code: 13394).

Data availability

Data, data dictionaries, analysis script and materials related to this study are publicly available on the Open Science Framework at https://osf.io/hpsq4. The study protocol and materials were registered on 06 February 2023 at https://osf.io/5h7bu. Discrepancies between this manuscript and the registered protocol are outlined in the Supplementary Material A. To facilitate reproducibility, this results section of the manuscript was written by interleaving regular prose and analysis code using R Markdown. The relevant files are available in a Code Ocean container (https://doi.org/10.24433/CO.0483372.v1) which recreates the software environment in which the original analyses were performed. This container allows the results section of the manuscript to be reproduced from the data and code with a single button press.

Contributions

Conceptualization:	Robert T. Thibault.
Data curation:	Robert T. Thibault and Deborah Bailey-Rodriguez.
Formal analysis:	Robert T. Thibault and Deborah Bailey-Rodriguez.
Funding acquisition:	Robert T. Thibault.
Investigation:	Robert T. Thibault, Deborah Bailey-Rodriguez, James E. Bartlett, Paul Blazey, Robin J. Green, and Madeleine Pownall.
Methodology:	Robert T. Thibault, Deborah Bailey-Rodriguez, James E. Bartlett, Paul Blazey, Robin J. Green, and Madeleine Pownall.
Project administration:	Robert T. Thibault.
Resources:	Robert T. Thibault.
Software:	Robert T. Thibault and Paul Blazey.
Supervision:	Robert T. Thibault and Marcus R. Munafò.
Validation:	Robert T. Thibault and James E. Bartlett.
Visualization:	Robert T. Thibault.
Writing - original draft:	Robert T. Thibault.

Writing - review & editing: Robert T. Thibault, Deborah Bailey-Rodriguez, James E. Bartlett, Paul Blazey, Robin J. Green, Madeleine Pownall, and Marcus R. Munafò.

Funding

Robert Thibault is supported by a general support grant awarded to METRICS from Arnold Ventures and a postdoctoral fellowship from the Canadian Institutes of Health Research. This project did not receive specific funding. This publication is the work of the authors and Robert Thibault will serve as guarantor for the contents of this paper. The funders have no role in the preparation of this manuscript or the decision to publish.

Competing interests

Robin J. Green is a member of the BPS Undergraduate Education Committee, who were tasked with updating the accreditation standards. All other authors declare no competing interests.

Acknowledgements

We thank Patricia Lyons from the British Psychological Society for collaborating in the development of this project. We thank Charlotte Pennington and Patricia Lyons for feedback on drafts of the initial set of Delphi items.

Endorsing organisations

We collaborated with three members of the BPS to run this study. The BPS Accreditation Operations Manager (Patricia Lyons) and Chair of the BPS Undergraduate Education Committee (Simon Goodson) helped design the survey. The co-lead of the Research Methods Working Group part of the Undergraduate Education Committee (Robin Green) helped design the study, execute the study, and write this manuscript. We asked several organisations if they could advertise this Delphi study by sharing it with their contacts (e.g., UK Reproducibility Network, ReproducibiliTea, Framework for Open and Reproducible Research Training—FORRT). This is the only capacity in which these organisations were involved.

References

Aiken, L. S., West, S. G., & Millsap, R. E. (2008). Doctoral training in statistics, measurement, and methodology in psychology: Replication and extension of Aiken, West, Sechrest, and Reno's (1990) survey of PhD programs in North America. *American Psychologist*, *63*(1), 32–50. https://doi.org/10.1037/0003-066X.63.1.32

Aiken, L. S., West, S. G., Sechrest, L., & Reno, R. R. (1990). Graduate Training in Statistics,

Methodology, and Measurement in Psychology. *American Psychologist*, 14.

Anglin, S. M., & Edlund, J. E. (2020). Perceived need for reform in field-wide methods and the teaching of replication, interpretation, and transparency. *Psychology Learning and Teaching*, *19*(1), 60–76. https://doi.org/10.1177/1475725719859453

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

British Psychological Society. (2019). *Standards for the accreditation of undergraduate, conversion and integrated Masters programmes in psychology*.

Cumming, G., & Calin-Jageman, R. (2016). *Introduction to the New Statistics | Estimation, Open Science, and Beyond*. https://www-taylorfrancis-com.stanford.idm.oclc.org/books/mono/10.4324/9781315708607/introduction-new-statistics-geoff-cumming-robert-calin-jageman

Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*, *67*(4), 401–409. https://doi.org/10.1016/j.jclinepi.2013.12.002

Field, A. P. (2014). *Skills in Mathematics and Statistics in Psychology and tackling transition*. https://s3.eu-west-2.amazonaws.com/assets.creode.advancehe-document-manager/documents/hea/private/resources/tt_maths_psychology_1568037242.pdf

Friedrich, J., Buday, E., & Kerr, D. (2000). Statistical training in psychology: A national survey and commentary on undergraduate programs. *Teaching of Psychology*, *27*(4), 248–257. https://doi.org/10.1207/S15328023TOP2704_02

Friedrich, J., Childress, J., & Cheng, D. (2018). Replicating a national survey on statistical training in undergraduate psychology programs: Are there “New statistics” in the new millennium? *Teaching of Psychology*, *45*(4), 312–323. https://doi.org/10.1177/0098628318796414

Gibson, S., & Sullivan, C. (2018). A changing culture? Qualitative methods teaching in U.K. psychology. *Qualitative Psychology*, *5*(2), 197–206. https://doi.org/10.1037/qup0000100

Hugh-Jones, S., Madill, A., Gibson, S., Keane, K., & Beestin, L. (2012). *A national survey of qualitative research methods teaching in UK HE psychology departments.* https://www.academia.edu/8483353/Hugh-Jones_S._Madill_A._et_al._2012_._A

Humphrey-Murto, S., Crew, R., Shea, B., Bartlett, S. J., March, L., Tugwell, P., Maxwell, L. J.,

Beaton, D., Grosskleg, S., & de Wit, M. (2019). Consensus Building in OMERACT: Recommendations for Use of the Delphi for Core Outcome Set Development. *The Journal of Rheumatology*, *46*(8), 1041–1046. https://doi.org/10.3899/jrheum.181094

Humphrey-Murto, S., Varpio, L., Wood, T. J., Gonsalves, C., Ufholz, L.-A., Mascioli, K., Wang, C., & Foth, T. (2017). The Use of the Delphi and Other Consensus Group Methods in Medical Education Research: A Review. *Academic Medicine: Journal of the Association of American Medical Colleges*, *92*(10), 1491–1498. https://doi.org/10.1097/ACM.0000000000001812

McAleer, P., Stack, N., Woods, H., DeBruine, L., Paterson, H., Nordmann, E., Kuepper-Tetzel, C. E., & Barr, D. J. (2022). *Embedding Data Skills in Research Methods Education: Preparing Students for Reproducible Research*. PsyArXiv. https://doi.org/10.31234/osf.io/hq68s

Palmer, W., Schlepper, L., Hemmings, N., & Crellin, N. (2021). *The right track: Participation and progression in psychology career paths*. https://www.nuffieldtrust.org.uk/research/the-right-track-participation-and-progression-in-psychology-career-paths

Pennington, C. (2023). *A Student's Guide to Open Science: Using the Replication Crisis to Reform Psychology*. Open University Press. https://www.mheducation.co.uk/a-student-s-guide-to-open-science-using-the-replication-crisis-to-reform-psychology-9780335251162-emea-group

Pownall, M., Azevedo, F., König, L. M., Slack, H. R., Evans, T. R., Flack, Z., Grinschgl, S., Elsherif, M. M., Gilligan-Lee, K. A., de Oliveira, C. M. F., Gjoneska, B., Kalandadze, T., Button, K., Ashcroft-Jones, S., Terry, J., Albayrak-Aydemir, N., Děchtěrenko, F., Alzahawi, S., Baker, B. J., … null, null. (2023). Teaching open and reproducible scholarship: A critical review of the evidence base for current pedagogical methods and their outcomes. *Royal Society Open Science*, *10*(5), 221255. https://doi.org/10.1098/rsos.221255

Spiegelhalter, D. (2019). *The Art of Statistics: How to Learn from Data*. https://www.amazon.com/Art-Statistics-How-Learn-Data/dp/1541618513

TARG Meta-Research Group. (2022). *Statistics education in undergraduate psychology: A survey of UK curricula*. Collabra: Psychology. https://doi.org/10.31234/osf.io/jv8x3

The British Academy. (2015). *COUNT US IN QUANTITATIVE SKILLS FOR A NEW GENERATION*. https://www.thebritishacademy.ac.uk/documents/220/Count-Us-In.pdf

Wiggins, S., Gordon-Finlayson, A., Becker, S., & Sullivan, C. (2016). Qualitative undergraduate project supervision in psychology: Current practices and support needs of supervisors

across North East England and Scotland. *Qualitative Research in Psychology*, *13*(1), 1–19. https://doi.org/10.1080/14780887.2015.1075641

Supplementary material A. Deviations from the preregistered protocol

- Whereas the protocol stated “All participants who completed Round 1 will be invited via email to participate in Round 2”, the manuscript states **“**All participants who began Round 1 were invited via email to participate in Round 2.” We made this decision after Round 1 was complete. The goal was to increase the number of respondents in Round 2, even if they didn’t complete Round 1. Nonetheless, the final dataset only includes participants who completed both rounds. We used data from participants who didn’t complete either Round 1 or Round 2 for a sensitivity analysis.

- The protocol states “We will advertise this Delphi study via mailing lists, psychology related websites, and social media (e.g., Twitter, Mastodon, LinkedIn)”. In the end, we were not successful in getting psychology organisations to post our study on their website. Our study team did not personally advertise the study on Mastodon or Linkedin, although others may have.

- The protocol did not specify how we planned to analyse the textual data. This was decided after data collection was complete.

- The protocol did not specify how we would summarise the items reaching consensus or how we would go from the results to our recommendations. These decisions were made after viewing the data. The results informed these processes.

- The protocol states “We will also perform a sensitivity analysis by excluding participants who failed the attention check question.” Instead, we decided to exclude these participants from the main analysis and then perform a sensitivity analysis with these participants. This decision was made before viewing the data.

- We report several instances of quantitative data that were not specifically prespecified, including Figure 2 (flowchart), Table 1 (participant characteristics), Table 3 (sensitivity analyses), Supplementary Table 1, and Supplementary Table 2.

Supplementary Materials B. Survey home page (incl. literature review and consent form)

[Participants viewed the text below before starting Round 1 of the Delphi study.]

Welcome to the **Survey on Research Methods Education in Undergraduate Psychology Programmes**, run in partnership with the British Psychological Society.

Before beginning the survey, please take a moment to read the following sections of this page:

1. Study summary
2. Stakeholder groups
3. Literature review
4. Consent form / participant information sheet

1. Study Summary

This study asks members of the UK psychology community about the research methods skills they believe psychology undergraduates should learn. The British Psychological Society (BPS) has partnered with us to run this study and they plan to use the results to inform the upcoming version of their accreditation standards.

This is a two-round survey (a Delphi study). The first round is similar to a normal survey. However, you will be able to comment on each question and suggest additional questions. Approximately a week after the first round closes, you will be invited to participate in the second round. In round 2, you will be shown the responses of other participants who completed round 1 and will be asked to re-rate some questions. A few questions will be added or modified based on comments from round 1.

2. Stakeholder groups

On the next page, you will be asked to select which stakeholder group you are part of (according to the instructions below):

A. Student in psychology (or recent graduate)
B. Research methods instructor in psychology
C. Academic based in psychology
D. Non-academic working in psychology

If you are an undergraduate or graduate student in psychology, please select option 'A' regardless of whether you teach or do research. If you are no longer a student, do not fit into

any of the other 3 stakeholder groups, and completed your undergraduate degree less than 3 years ago, select option 'A'.

If you teach or coordinate psychology research methods in an undergraduate or Masters conversion programme, and are not a student, please select option 'B', regardless of whether you are also an academic or non-academic psychologist.

If you are both an academic and non-academic psychologist (e.g., work for industry and a university), select the group that you feel represents you best.

If you do not fit into one of these four stakeholder groups, you are ineligible to participate in this survey—in this case, please do not proceed.

3. Literature review

Why run this study?

To help research methods education keep abreast with advances in research practice and education. The results will inform the upcoming British Psychological Society's 'Standards for the accreditation of undergraduate, conversion and integrated Masters programmes in psychology'.

This review provides a quick summary on research methods education in UK undergraduate psychology programmes. More detailed reports are available, here, here, and here. A list of relevant articles appears in the references at the bottom of this page.

There are no correct or incorrect answers to this survey. This summary is to help orient you as a survey participant. Please provide answers based on your own thoughts.

Statistics in psychology research

Statistical shortcomings and errors are common in psychology research[e.g.,1]. They can increase the prevalence of biased results and lead to a distorted and uncertain evidence base[e.g.,2]. A 2015 British Academy report[3] stated that "A co-ordinated and continuous effort at improving quantitative skills across all phases of education and employment…is therefore now urgently needed."

Quantitative research methods education

Only a few articles describe the quantitative skills taught in undergraduate psychology programmes. They find minimal changes to the content of undergraduate statistics modules in the United States over 20 recent years[4,5] and that the emphasis psychology programmes place

on null hypothesis significance testing appears out of step with modern statistical thought in the psychological sciences (e.g., estimation, uncertainty, and open science)[5–8].

Open Science practices are not yet widespread across psychology education. Nonetheless, relevant teaching resources exist[e.g.,9,10] and incoming students support Open Science norms[11].

As of 2019, few UK undergraduate psychology programmes have publicly available syllabi for their quantitative research methods modules.[8]

Software and resources

Of 27 UK undergraduate psychology programmes assessed in 2019, 26 use SPSS, 8 use Excel, and 1 to 3 use R, depending on whether option modules are included.[8] Andy Field's *Discovering Statistics using IBM SPSS Statistics*[12] was the most commonly used textbook. Field has written an equivalent book for R[13] and open source online textbooks exist (e.g., PsyTeachR).

Qualitative research methods education

Although qualitative research skills have been part of the BPS accreditation standards since 2004, there is limited research on how they are taught.[14,15] Data suggests that some instructors and students see qualitative research as inferior to quantitative research[14,16] and that additional training and resources are needed to effectively teach qualitative methods.[15,17]

There is currently no established consensus on how qualitative approaches are taught in UK psychology programmes. This can present difficulties, but also allows for instructors to teach to their strengths.[18]

Teaching and evaluation formats

Of 27 UK psychology programmes, many research methods modules mentioned dedicating hours to workshop, labs, and practicals. However, none mentioned teaching formats such as a flipped classroom, resequenced content, or collaborative learning. While free online modules gain popularity (e.g., Improving your statistical inferences) it remains unclear whether undergraduate instructors are leveraging these resources.

Research on quantitative education suggests that using a problem solving framework and collaborative learning environment can reduce anxiety around statistics and improve learning outcomes.[19,20] Both instructors and students identify statistics anxiety as a key factor inhibiting their development of quantitative skills.[21] Several modules, however, grade 100% based on exams[8], which can induce anxiety and fail to mimic the contexts in which students will use quantitative skills in their future.

Policies and guidelines

The BPS provides accreditation standards for undergraduate programmes, which includes a page on research methods. They also provide a 24-page supplementary guidance for teaching

research methods. The national bodies in other countries we checked (e.g., United States, Canada) do not accredit psychology programmes.

The accreditation standards consist of broad statements such as “students should be able to analyse, present and evaluate quantitative and qualitative data”. It is difficult to assess whether such standards are being met.

The present study aims to provide more detailed and clear-cut information to help update the BPS accreditation standards.

References

1. Nuijten, M. B., Hartgerink, C. H. J., van Assen, M. A. L. M., Epskamp, S. & Wicherts, J. M. The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods* **48**, 1205–1226 (2016).

2. Simmons, J. P., Nelson, L. D. & Simonsohn, U. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* **22**, 1359–1366 (2011).

3. The British Academy. *COUNT US IN QUANTITATIVE SKILLS FOR A NEW GENERATION*. https://www.thebritishacademy.ac.uk/documents/220/Count-Us-In.pdf (2015).

4. Friedrich, J., Buday, E. & Kerr, D. Statistical training in psychology: A national survey and commentary on undergraduate programs. *Teaching of Psychology* **27**, 248–257 (2000).

5. Friedrich, J., Childress, J. & Cheng, D. Replicating a national survey on statistical training in undergraduate psychology programs: Are there “New statistics” in the new millennium? *Teaching of Psychology* **45**, 312–323 (2018).

6. Calin-Jageman, R. J. & Cumming, G. The New Statistics for Better Science: Ask How Much, How Uncertain, and What Else Is Known. *The American Statistician* **73**, 271–280 (2019).

7. Anglin, S. M. & Edlund, J. E. Perceived need for reform in field-wide methods and the teaching of replication, interpretation, and transparency. *Psychology Learning and Teaching* **19**, 60–76 (2020).

8. TARG Meta-Research Group. *Statistics education in undergraduate psychology: A survey of UK curricula*. https://osf.io/jv8x3 (2022) doi:10.31234/osf.io/jv8x3.

9. Pownall, M. *et al.* Embedding open and reproducible science into teaching: A bank of lesson plans and resources. *Scholarship of Teaching and Learning in Psychology* No Pagination Specified-No Pagination Specified (2021) doi:10.1037/stl0000307.

10. Pennington, C. *A Student's Guide to Open Science: Using the Replication Crisis to Reform Psychology*. (Open University Press, 2023).

11. Beaudry, J. L., Williams, M., Philipp, M. C. & Kothe, E. What do incoming university students believe about open science practices in psychology? Preprint at https://doi.org/10.31234/osf.io/rq3zs (2021).

12. Field, A. Discovering statistics using IBM SPSS statistics. in *Statistics* (2013).

13. Field, A., Miles, J. & Field, Z. *Discovering Statistics Using R*. (SAGE Publications Ltd, 2012).

14. Gibson, S. & Sullivan, C. A changing culture? Qualitative methods teaching in U.K. psychology. *Qualitative Psychology* **5**, 197–206 (2018).

15. Hugh-Jones, S., Madill, A., Gibson, S., Keane, K. & Beestin, L. A national survey of qualitative research methods teaching in UK HE psychology departments. (2012).

16. Clarke, V. & Braun, V. *Methods: Teaching thematic analysis*. https://www.bps.org.uk/volume-26/edition-2/methods-teaching-thematic-analysis (2013).

17. Wiggins, S., Gordon-Finlayson, A., Becker, S. & Sullivan, C. Qualitative undergraduate project supervision in psychology: current practices and support needs of supervisors across North East England and Scotland. *Qualitative Research in Psychology* **13**, 1–19 (2016).

18. Gibson, S. & Sullivan, C. Teaching Qualitative Research Methods in Psychology: An Introduction to the Special Issue. *Psychology Learning & Teaching* **11**, 1–5 (2012).

19. Spiegelhalter, D. *The Art of Statistics: How to Learn from Data*. (2019).

20. Research Articles. https://www.youcubed.org/evidence/research-articles/.

21. Field, A. P. *Skills in Mathematics and Statistics in Psychology and tackling transition*. https://s3.eu-west-2.amazonaws.com/assets.creode.advancehe-document-manager/documents/hea/private/resources/tt_maths_psychology_1568037242.pdf (2014).

4. Consent form / Participant information sheet

Thank you for your interest in completing this anonymous Delphi survey. Please take time to read the following information carefully and discuss it with others if you wish. Your participation is voluntary.

What is the purpose of the research?
We are interested in learning what the psychology community in the UK thinks that undergraduate psychology students should learn in terms of research methods. More specifically, we are running this study in partnership with the British Psychological Society (BPS)

and they plan to use the results to inform their upcoming accreditation standards for undergraduate psychology education. We are inviting any psychologist, psychology instructor, or psychology student in the UK, or with a link to the UK psychology community, to participate.

How much time will the study take?
This is a two-round survey. The first round should take approximately 20-30 minutes, depending on whether you provide comments on the questions. It will be open for two weeks, and you can save your progress and return to the survey anytime in these two weeks.

The second round will begin approximately one week after the first round is completed. It should take approximately 20-30 minutes, will be open for two weeks, and progress can be saved.

What will happen to the results of this questionnaire?
When the project is finished, we will share the data with the British Psychological Society. We will analyse the data and report the findings. We may report this in a scientific journal and/or present at a scientific meeting. If you would like a copy of the final paper, you may request this by contacting robert.thibault@bristol.ac.uk. Data will be treated in the following ways:

What will happen to my data?
Your involvement in the study will remain confidential. This information will only be available to research staff and national bodies which monitor whether research studies are conducted properly. Your study data will be anonymised. This means that it will be given an identification number and any identifying information about you will be removed.

(1) Answers to multiple-choice questions will be stored as open data on the Open Science Framework (osf.io) and the University of Bristol Research Data Repository. They will not include identifying information. Open data are made available, free of charge, to anyone interested in the project, or who wishes to conduct their own analyses of the data.

(2) Comments and open-ended questions will also be stored as open data on the Open Science Framework and the University of Bristol Research Data Repository. Parts of responses may be redacted to maintain anonymity. Although these data are anonymous and will contain redactions of information that may compromise anonymity, depending on the content of your response, they may still be identifiable.

Why open data?
Open access to research findings and access to data is considered best research practice and is a requirement of many funding bodies and journals. As a large proportion of research is publicly funded, the outcomes of the research should be made publicly available. Sharing data helps to

maximise the impact of investment through wider use, and encourages new avenues of research.

Can I withdraw my data after I have taken part?
Yes, before the data are made open. Although the study team will not be able to identify which data are yours, the DelphiManager team will be able to associate a unique identifier with your email address and your responses. This information is necessary for a Delphi study because respondents need to be reminded of their round 1 responses, when responding to round 2. To withdraw your data, email the study lead at robert.thibault@brsitol.ac.uk. After the study is complete and the data are made open, you will no longer be able to withdraw your data.

Who has reviewed/approved this work?
This project has received ethics approval from the University's Faculty of Life Sciences Research Ethics Committee at the University of Bristol [ENTER ID WHEN ACCEPTED]. If you have any concerns related to this project, please direct them to the Psychological Science Human Research Ethics Committee, via Liam McKervey (liam.mckervey@bristol.ac.uk or +44 (0)117 928 7841).

Contact details
If you have any questions about the study, please contact Robert Thibault (robert.thibault@bristol.ac.uk) at the School of Psychological Science, University of Bristol.

I understand that after the study the data will be made "open data". I understand that this means the anonymised data will be publicly available and may be used for purposes not related to this study, and it is unlikely that someone will be able to identify me from these data.

By clicking the "Register" button below, you are providing your informed consent to participate in this study.

[Participants viewed the text below before starting Round 1 of the Delphi study.]

Questions - Round 2

Please rate the question below.

The figure presents the distribution of responses to this question from each stakeholder group in Round 1. The figures are presented to provide you with information about the perspectives of each stakeholder group.

This information may or may not impact your rating. Please note that the barcharts are presented as percentages and each group had a different number of respondents. The column "X" in each barchart depicts the percentage of respondents who selected "Unable to rate".

Respondents in the Research Methods Instructor group, may also be academic or non-academic psychologists. However, their responses are only displayed in the Research Methods Instructor figure. Respondents in the Academic Psychologist and Non-Academic Psychologist groups, are not research methods instructors. To see the full definition for each stakeholder group, click on the 'About' tab in the bottom left corner of the page on scroll to section 2. Stakeholder Groups.

Supplementary Material C. Participant invitation templates

We provided organisations interested in advertising our study with the following invitation templates. The decision on the exact invitation text to use will be left to each individual organisation to decide.

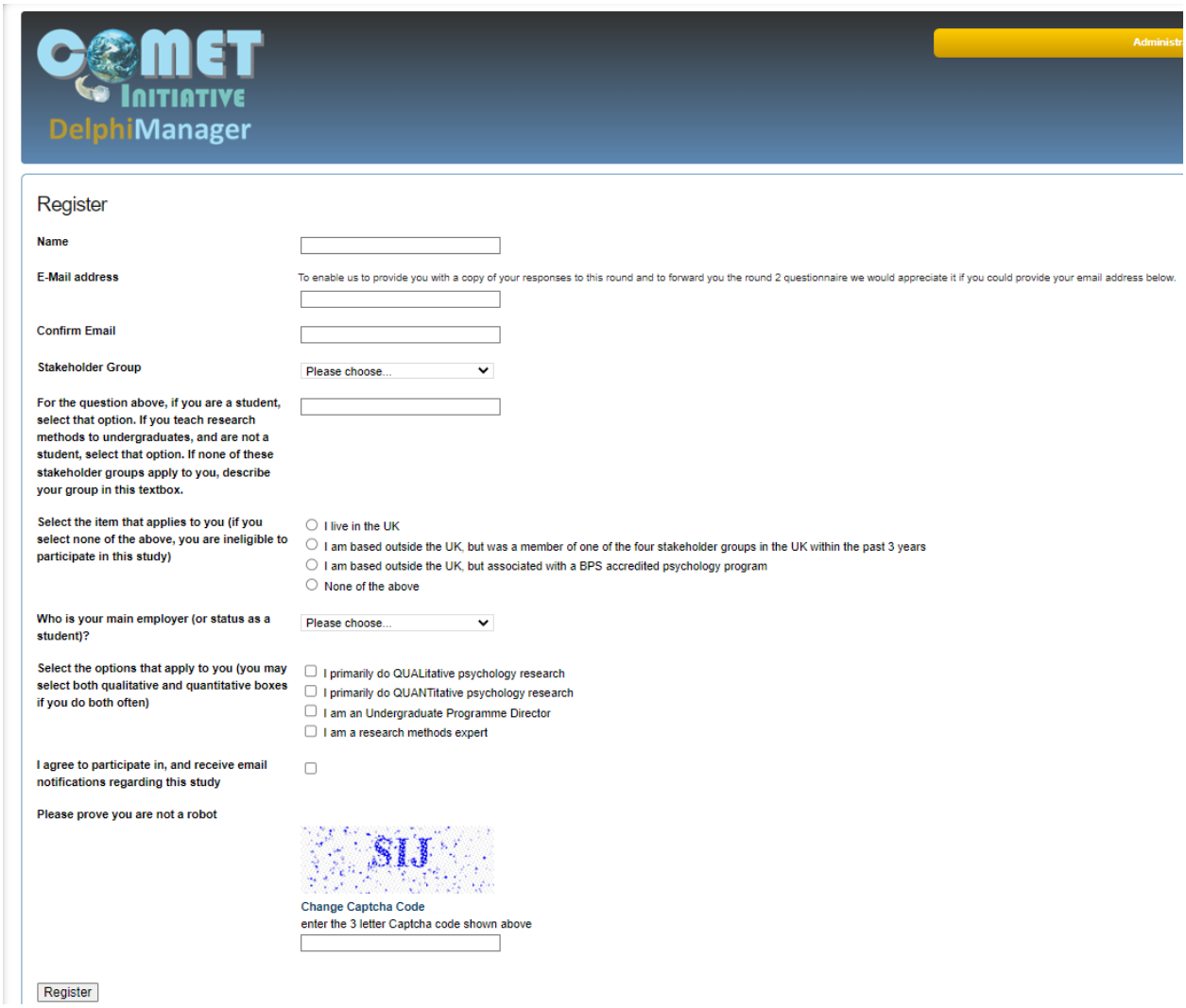
Short version (for social media)

> Have thoughts about what psychology undergrads should learn? Then participate in this Delphi study run in partnership with the British Psychological Society [LINK TO DELPHI]. Your input will shape the future of quantitative and qualitative research methods education in the UK.

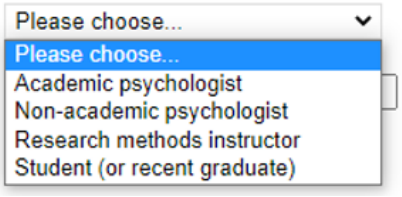
Long version (for email)

> Dear [colleagues],
>
> We would like to invite you to participate in a study about undergraduate psychology education in the UK [LINK TO DELPHI].
>
> This study aims to reach a consensus on the qualitative and quantitative research methods that UK psychology undergraduates should learn. The study is run in partnership with the British Psychological Society (BPS) and they will use the results to inform the upcoming version of their accreditation standards for undergraduate programmes.
>
> This study uses the Delphi technique, which systematically elicits two-rounds of anonymous and asynchronous input from a range of stakeholders. For more details, please follow the link to the study.
>
> We invite instructors, academic psychologists, non-academic psychologists, and psychology students to participate.
>
> If you have any questions or comments, please send them directly to the study lead at robert.thibault@bristol.ac.uk
>
> Kind regards,
> [signature]

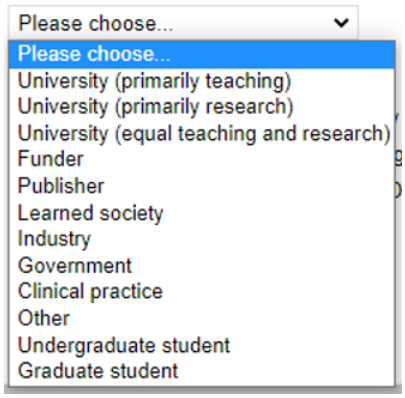
Supplementary Materials D. Study registration form



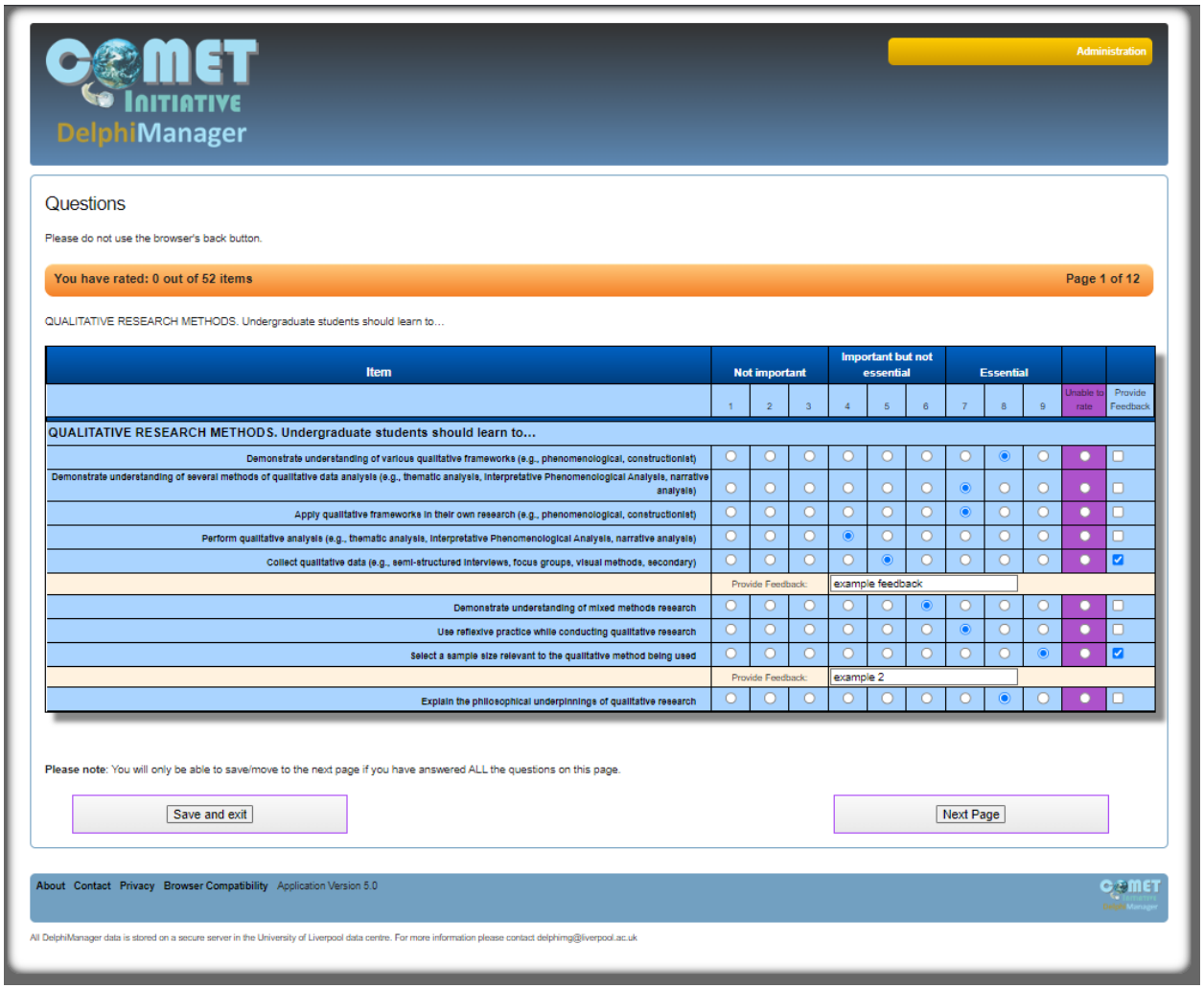
Drop-down menu 1



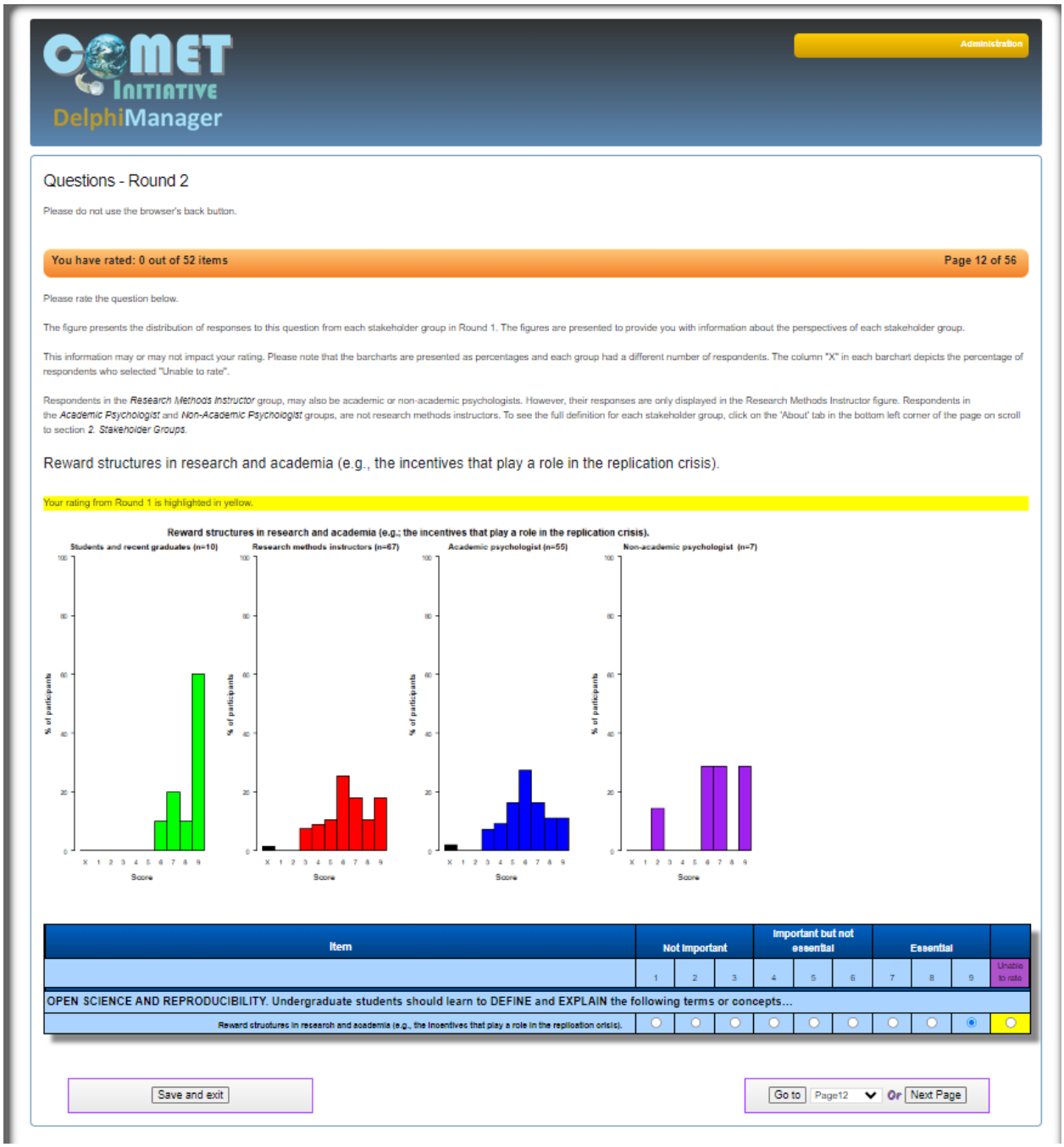
Drop-down menu 2


Please choose...
Please choose...
University (primarily teaching)
University (primarily research)
University (equal teaching and research)
Funder
Publisher
Learned society
Industry
Government
Clinical practice
Other
Undergraduate student
Graduate student

Supplementary Materials E. Additional figures and tables



Supplementary Figure 1. Screenshot of DelphiManager Round 1. Note, this screenshot was taken after the study was complete, and thus the text "0 out of 52 items" appears, whereas "0 out of 72" appeared for participants.



Supplementary Figure 2. Screenshot of DelphiManager Round 2. In Round 2, items were presented one at a time alongside feedback from participants' ratings from the previous round.

Supplementary Table 1. Counts for open-ended responses. Few participants accounted for a large proportion of the feedback. For example, participant BPSRE00038 provided 15% (33/222) of all open-ended feedback on specific Delphi items. We considered this uneven distribution of open-ended responses when analysing this data. Participants could only leave a single general feedback comment at the very end of each round. Based on the format in which the DelphiManager outputs data, the user_ids from the column 'items suggested to include' cannot be linked to the user_ids in the other columns.

Feedback on items		Items suggested to include		Reasoning for changing a rating across a boundary	
user_id	frequency	user_id	frequency	user_id	frequency
BPSRE00038	33	22	28	BPSRE00015	14
BPSRE00005	22	5	5	BPSRE00156	14
BPSRE00087	15	17	4	BPSRE00025	12
BPSRE00029	14	9	3	BPSRE00054	11
BPSRE00026	11	10	3	BPSRE00084	10
BPSRE00076	11	11	3	BPSRE00164	10
BPSRE00074	10	12	3	BPSRE00082	9
BPSRE00052	9	13	3	BPSRE00107	9
BPSRE00077	9	23	3	BPSRE00018	8
BPSRE00009	8	2	2	BPSRE00067	8
BPSRE00120	8	3	2	BPSRE00074	8
BPSRE00053	7	16	2	BPSRE00160	8
BPSRE00044	6	20	2	BPSRE00030	7
BPSRE00124	5	21	2	BPSRE00055	7
BPSRE00081	4	28	2	BPSRE00094	7
BPSRE00110	4	1	1	BPSRE00130	7
BPSRE00150	4	4	1	BPSRE00145	7
BPSRE00154	4	6	1	BPSRE00019	6
BPSRE00075	3	7	1	BPSRE00022	6
BPSRE00079	3	8	1	BPSRE00034	6
BPSRE00167	3	14	1	BPSRE00052	6
BPSRE00047	2	15	1	BPSRE00076	6
BPSRE00050	2	18	1	BPSRE00088	6
BPSRE00057	2	19	1	BPSRE00089	6
BPSRE00082	2	24	1	BPSRE00151	6
BPSRE00089	2	25	1	BPSRE00005	5
BPSRE00091	2	26	1	BPSRE00008	5
BPSRE00106	2	27	1	BPSRE00077	5
BPSRE00139	2	29	1	BPSRE00108	5
BPSRE00152	2	30	1	BPSRE00115	5
BPSRE00008	1			BPSRE00137	5
BPSRE00025	1			BPSRE00147	5
BPSRE00034	1			BPSRE00154	5
BPSRE00035	1			BPSRE00172	5

Feedback on items		Items suggested to include		Reasoning for changing a rating across a boundary	
user_id	frequency	user_id	frequency	user_id	frequency
BPSRE00042	1			BPSRE00010	4
BPSRE00065	1			BPSRE00042	4
BPSRE00094	1			BPSRE00091	4
BPSRE00098	1			BPSRE00092	4
BPSRE00099	1			BPSRE00136	4
BPSRE00162	1			BPSRE00011	3
BPSRE00166	1			BPSRE00013	3
				BPSRE00017	3
				BPSRE00024	3
				BPSRE00026	3
				BPSRE00038	3
				BPSRE00127	3
				BPSRE00131	3
				BPSRE00141	3
				BPSRE00142	3
				BPSRE00144	3
				BPSRE00152	3
				BPSRE00016	2
				BPSRE00023	2
				BPSRE00028	2
				BPSRE00057	2
				BPSRE00066	2
				BPSRE00106	2
				BPSRE00117	2
				BPSRE00124	2
				BPSRE00138	2
				BPSRE00162	2
				BPSRE00029	1
				BPSRE00070	1
				BPSRE00073	1
				BPSRE00075	1
				BPSRE00111	1
				BPSRE00118	1
				BPSRE00120	1
				BPSRE00126	1
				BPSRE00133	1
				BPSRE00149	1
				BPSRE00169	1
				BPSRE00180	1

Supplementary Table 2. Consensus results sorted by the block (domain) in which the item was presented.

Domain	Items reaching consensus	Items total	% reaching consensus
stats	6	8	75
design	10	15	67
data	4	8	50
quant	4	8	50
OS	6	15	40
misc	3	8	38
qual	1	10	10
resources	0	6	0

Supplementary Materials F. Delphi items

- Sections were presented in a random order.
- Questions within each section always appeared in the same order (the DelphiManager Software does not allow randomization of the questions within a section).
- Each question…
    - is rated on a scale from 1-9, where 1-3 is 'not important' 4-6 is 'important, but not essential', and 7-9 is 'essential'.
    - has the option to select 'unable to rate'
    - has the option to provide open-ended written feedback

Number of items to rate in Round 1:	72
Number of items removed after Round 1:	27
Number of items added after Round 1:	7
Number of items in Round 2:	52

---

1. **STATISTICAL ANALYSES.** *Undergraduate students should learn how to <u>calculate/perform</u>…*
    - 1.1. Descriptive statistics (e.g., mean, interquartile range)
    - 1.2. Significance tests (e.g., t-test, ANOVAs, Pearson's correlations, chi-squared)
    - 1.3. Parameter estimation (e.g., calculating 95% confidence intervals)
    - 1.4. Regression
    - 1.5. Methods to assess statistical assumptions (e.g., normality)
    - 1.6. Effect sizes (e.g., Cohen's d, odds ratios)
    - 1.7. Equivalence testing
    - 1.8. Factor analysis

2. **QUANTITATIVE DATA SKILLS.** *Undergraduate students should learn how to…*
    - 2.1. Identify and categorise different types of data (e.g., binary, continuous, categorical)
    - 2.2. Clean data (e.g., remove bad data, rearrange data—i.e., "wrangle" data)
    - 2.3. Represent data visually (e.g., create histograms, line graphs)
    - 2.4. Use a programming language to manage and analyse data (e.g., R or Python)
    - 2.5. Use a statistical analysis package with a graphical user interface (e.g., SPSS or JASP)
    - 2.6. Simulate data
    - 2.7. Use descriptive statistics effectively before learning to perform inferential statistical tests

3. **QUANTITATIVE RESEARCH METHODS CONCEPTS.** Undergraduate *students should learn to <u>define and explain</u> the following concepts…*

3.1. The existence of different statistical frameworks, including frequentist statistics and Bayesian statistics
3.2. The existence of different statistical approaches, including parameter estimation and significance testing
3.3. The importance of descriptive statistics and how they differ from inferential statistics
3.4. The difference between statistical significance and practical significance (also called clinical, theoretical, or biological significance)
3.5. The value of exploratory research and how it differs from confirmatory research
3.6. Probability and randomness
3.7. Psychometrics (e.g., scale construction)

4. QUALITATIVE RESEARCH METHODS. Undergraduate *students should learn to…*

4.1. Demonstrate understanding of various qualitative frameworks (e.g., phenomenological, constructionist)
4.2. Demonstrate understanding of several methods of qualitative data analysis (e.g., thematic analysis, Interpretative Phenomenological Analysis, narrative analysis)
4.3. Apply qualitative frameworks in their own research (e.g., phenomenological, constructionist)
4.4. Perform qualitative analysis (e.g., thematic analysis, Interpretative Phenomenological Analysis, narrative analysis)
4.5. Collect qualitative data (e.g., semi-structured interviews, focus groups, visual methods, secondary)
4.6. Demonstrate understanding of mixed methods research
4.7. Use reflexive practice while conducting qualitative research
4.8. Critically appraise qualitative research (e.g., using qualitative criteria)
4.9. Select a sample size relevant to the qualitative method being used
4.10. Explain the philosophical underpinnings of qualitative research

5. RESEARCH DESIGN. Undergraduate *students should learn how to…*

5.1. Formulate a research question
5.2. Explain the difference between a research question and a hypothesis
5.3. Design a study to answer a specific research question (including the selection of an appropriate research method and analytic approach)
5.4. Create a sampling plan and data collection plan (for both quantitative and qualitative research; and in line with the research question and method of analysis)
5.5. Determine a smallest effect size of interest (SESOI) for quantitative research.
5.6. Perform sample size calculations for quantitative research (e.g., power calculations, precision calculations)
5.7. Attention check: Please select “3” to confirm you are reading these questions.
5.8. Assess validity and reliability
5.9. Operationalize all elements of a study
5.10. Apply experimental and non-experimental research designs

5.11. Apply blinding and randomization when conducting an experiment
5.12. Follow accepted reporting guidelines
5.13. Design a survey
5.14. Identify and assess ethical issues (in both qualitative and quantitative research)

6. OPEN SCIENCE AND REPRODUCIBILITY. Undergraduate *students should learn to define and explain the following terms or concepts……*

6.1. The "replication crisis"
6.2. Philosophy of science
6.3. Questionable Research Practices (QRPs) (e.g., selective reporting, p-hacking)
6.4. Research misconduct (i.e., fabrication, falsification, and plagiarism)
6.5. Replication studies and reproducibility
6.6. Generalisability and robustness
6.7. Reward structures in research and academia (e.g., the incentives that play a role in the replication crisis).
6.8. Preregistration and Registered Reports
6.9. Sources of bias (in both qualitative and quantitative research. E.g., sampling bias)
6.10. Cognitive biases (and how these drive the replication crisis. E.g., confirmation bias)
6.11. Data, code, and material sharing (e.g., open data).
6.12. The publication process (including Open Access and peer review)
6.13. Meta-research / meta-science (e.g., how these methods shed light on the replication crisis)
6.14. Multiverse analyses / many-analyst approaches
6.15. Systematic reviews and meta-analysis

7. ACCESSIBILITY OF RESOURCES. *Research methods modules in undergraduate psychology should…*

7.1. Only use freely available software (e.g., R or JASP, rather than SPSS)
7.2. Make their syllabi publicly available (e.g., on the Open Science Framework—OSF)
7.3. Provide students with syllabi that include a week-by-week outline of the module contents
7.4. Never be entirely graded with closed-book exams
7.5. Have a higher staff to student ratio than for non-research methods modules (this could include teaching assistants).

8. MISCELLANEOUS QUESTIONS

8.1. Students should preregister the quantitative aspects of their final year project.
8.2. Students should be allowed to perform a replication as their final year research project
8.3. Students should be allowed to conduct their final year project in a team.
8.4. Students should have the option to conduct a qualitative, quantitative, or mixed-methods project in their final year research.

8.5. Research methods instructors should be given time and support to improve their skills (e.g., to improve their understanding of qualitative methods or learn R, if they plan to teach these skills)

8.6. Research methods modules should actively employ teaching and grading methods known to reduce "statistics anxiety"

9. ITEMS ADDED TO ROUND 2

9.1. Undergraduate students should learn how to: anonymize data

9.2. Undergraduate students should learn how to: Consider diverse perspectives when designing a study (e.g.; global approaches; marginalised or vulnerable communities; decolonising methodologies)

9.3. Undergraduate students should learn how to: Demonstrate general computer skills for research (e.g.; file structure; version control; spreadsheets; and word processing)

9.4. Undergraduate students should learn how to: Identify basic study designs (e.g.; randomized trial; cross-sectional; qualitative designs)

9.5. Undergraduate Students should learn how to: Search and collate published research (e.g.; by using databases such as Scopus and reference managers such as Endnote or Zotero)

9.6. Undergraduate students should learn to DEFINE and EXPLAIN the following concepts: Alternative measures of effect sizes (e.g.; probability of superiority; Cohen's U3; number needed to treat)

9.7. Research methods MODULES in undergraduate psychology should: Emphasize skills that transfer beyond an academic research context (e.g.; training for the job market; creating an informed citizenry)

[The remainder of the questions were open-ended text responses]

10. ADDITIONAL OUTCOMES

10.1. If you feel that this survey did not include certain questions you would deem important, please enter them here and provide your ratings. We will review these questions and may include some in the next round of this Delphi study.

11. COMMENTS

11.1. Please provide any thoughts you have about this Delphi study here.

12. **RATING CHANGE QUESTION** (Round 2 only)

12.1. Some of the ratings you have changed have moved across the rating categories eg. from Not Important to Important but not critical. Please could you give a reason for these significant changes:

Supplementary Materials G. Recommendation to the BPS

We sent a synopsis of this Delphi study and a list of recommendations to the BPS Undergraduate Education Committee on 03 July 2023. This pdf document is available at (https://osf.io/9a6vx). The content of this pdf is also provided on the next 3 pages.

Recommendations for updating the research methods section of the BPS undergraduate accreditation standards

Authors. Robert T. Thibault, Deborah Bailey-Rodriguez, James Bartlett, Paul Blazey, Robin J. Green, Madeleine Pownall, Marcus R. Munafò. Correspondence to: robert.thibault@stanford.edu

Executive summary

Our team of researchers ran a consensus process. Our goal was to identify the research methods skills that the UK psychology community deems essential for undergraduates to learn. Of 78 items included in the consensus process, 34 reached consensus. We also performed a qualitative analysis of 707 open-ended text responses. Based on our findings, we developed nine Core Recommendations for updating the research methods section of the BPS accreditation standards. These include emphasising data skills, research design, descriptive statistics, critical analysis, qualitative methods, and both significance testing and parameter estimation; as well as giving precedence to foundational skills, promoting transferable skills, and creating space within curricula to enable these recommendations. A full account of the consensus process, including methods, results, and interpretation is available here [temporarylink to manuscript draft removed].

Methodology

We used the Delphi technique, which systematically elicits anonymous, asynchronous, and iterative input from a range of stakeholders. With input from the BPS Accreditation Operations Manager (Patricia Lyons) and Chair of the BPS Undergraduate Education Committee (Simon Goodson), a steering committee of 5 researchers and instructors developed a survey that consisted of 78 items for participants to rate as 'not important', 'important, but not essential', or 'essential' for undergraduate students in UK psychology programmes to learn. Invitations to participate were extended via email and social media. Anyone from the UK psychology community was welcome to participate. 103 research methods instructors, academic psychologists, non-academic psychologists, and students completed the consensus process. As is common in Delphi studies, we considered consensus reached if at least 75% of participants in each stakeholder group rated an item as 'essential'. Participants had several opportunities to provide open-ended feedback and we assessed the stability of responses with a second survey round. Similar consensus processes have been used to develop standards in over 200 medical education programmes (Humphrey-Murto et al., 2017).

Core Recommendations

We developed these recommendations based on ratings to the 78 Delphi items and participants' open-ended feedback. Each recommendation is presented as a general concept that the BPS could integrate into their accreditation standards, as well as specific text that could be added as a bullet point to *Section 2.1.4.g Research Methods*. Following the format of the current accreditations standards, the bullet points below are preceded by the text "Students should be able to…".

1. **Require a strong understanding of data and quantitative data skills.**
    - "Identify and categorise different types of quantitative data (e.g., categorical, continuous), clean and wrangle data, and represent data visually (e.g., histograms, line graphs)".
2. **Emphasise general skills in research design.**

- Edit the item "generate and explore hypotheses and research questions drawing on relevant theory and research" to "formulate and operationalise research questions and hypotheses, drawing on relevant theory and research; and explain the difference between a research question and hypothesis".

- "Identify basic study designs, differentiate exploratory and confirmatory research, and apply experimental and non-experimental research designs".

3. **Prioritise a solid foundation in descriptive statistics.**

    - "Calculate descriptive statistics (e.g.; mean; interquartile range), and explain how descriptive statistics differ from inferential statistics"

4. **Provide students with a framework to critically assess research claims.**

    - "Identify and explain sources of bias (e.g., sampling bias), cognitive biases (e.g., confirmation bias), questionable research practices (e.g., selective reporting), generalisability and robustness, research misconduct, replication studies and reproducibility".

5. **Raise the prominence of qualitative methods throughout the accreditation standards.**

    - "Understand several qualitative methods, perform qualitative analyses, present qualitative research, and critically appraise qualitative research (e.g. using qualitative criteria)."

6. **For inferential statistics, require that parameter estimation techniques, such as confidence intervals and effect sizes, be taught alongside significance testing.**

    - "Understand both parameter estimation and significance testing, calculate effect sizes and confidence intervals, and perform regressions".

    - "Explain the difference between statistical significance and practical significance".

7. **Give precedence to teaching foundational research methods skills (as outlined in Recommendations 1-5).**

    - Proficiently apply foundational qualitative and quantitative skills (e.g., descriptive statistics) before learning more advanced techniques (e.g., inferential statistics)".

8. **Promote content that elucidates how research methods skills transfer beyond academia.**

    - *Section 2.2 Teaching and learning* "Education providers must deliver content that demonstrates how students can apply research methods skills and psychological literacy in diverse real-world situations, as well as academic and non-academic employment pathways.

9. **Enable Recommendations 1-8 by encouraging research methods education throughout the programme, focusing on fewer skills in greater depth, and offering optional modules for more advanced research methods skills.**

- Change the statement: "Research methods must be delivered at Level 5 or Level 6" to "Research methods should be delivered throughout Levels 4-6 and can be integrated into core area modules".

- Edit the statement "carry out empirical studies involving a variety of methods of data collection, including experiments, observation, questionnaires, interviews and focus groups" to "carry out at least one quantitative and one qualitative study, which may include…"

- Edit the statement "use a variety of psychological tools, including specialist software, laboratory equipment and psychometric instruments" to "identify a variety of psychological tools, including…and use at least some of these".

Elaboration

Ratings were very high and consensus reached for data skills, basic research design, descriptive statistics, and inferential statistics. Almost 90% of participants rated it essential that students learn to use descriptive statistics effectively before learning to perform inferential statistics. Our qualitative analysis also highlighted the need for students to master foundational quantitative and qualitative skills, rather than attempt to perform analyses that they do not understand. These findings challenge the null-hypothesis significance testing (NHST) centric approach taken in many research methods curricula. The BPS can take a leadership role by de-emphasising NHST in favour of robust foundational quantitative and qualitative skills. Students must understand what analyses they are performing and why they are doing them.

Critical assessment is a pillar throughout the 2019 accreditation standards, but specific concepts and tools are not outlined. By updating the standards to include specific items, such as learning about replication and sources of bias, the BPS can champion a more structured approach to critically assess the psychology literature and other research claims.

Open-ended comments suggested that students should learn how to answer a research question and focus on fewer technical abilities. With this in mind, research methods education could adopt a problem-solving approach by teaching students how to ask a clear question, design an effective research plan, identify what data is needed to answer their question, and select an appropriate analysis plan or statistical procedure to answer their question. The BPS accreditation standards could outline this problem-solving approach as an alternative to statistics-centric methods curricula.

Additional considerations

44 items did not reach consensus. These items spanned topics including module format, final-year projects, computer skills, approaches to research, and advanced analysis techniques. Many of these items received a high-level of agreement, but fell short of consensus. There was **not** consensus *against* teaching these items. Notably, the survey was also designed with a quantitative focus.

Delphi studies are not designed to identify the best educational content or teaching methods. Instead, our results represent topics which a broad sample of the academic psychology community in the UK believe are essential for undergraduates to learn. This study measured community norms and expectations.