Special Issue: Metascience and Psychology Pedagogy

SOCIETY FOR THE TEACHING OF PSYCHOLOGY

Teaching of Psychology
1–9

DOI: 10.1177/00986283261451848
journals.sagepub.com/home/top
Sage

"Anyone Could Do This!": Multiverse as an Undergraduate Psychology Capstone Project

Peter J. Allen[1], Emily Abraham[1], Agnes Bleach[1], Rory Hill[1], Sana Mahmud[1], Jordan Wild[1], and Amanda E. Krause[2]

Abstract

Background: Heyman and Vanpaemel proposed "many-multiverses-one-dataset" as a pedagogically sound alternative to traditional undergraduate capstone projects. However, they did not evaluate this model. In our context, capstone projects require a higher level of individualization. Furthermore, project students value choice, autonomy, and ownership. Consequently, we have iteratively developed a variation of Heyman and Vanpaemel's model in which students choose their own effect/dataset for an individual multiverse capstone project.

Objective: To evaluate student-led individual multiverse capstone projects.

Method: One cohort of five project students qualitatively reflected on designing, running, and reporting an individual multiverse project in a small group supervision context. We thematically analyzed these reflections.

Findings: Two themes captured (a) the factors promoting project success (intrapersonal, interpersonal, and contextual) and (b) the outcomes of a multiverse capstone project (the range of skills developed and multiverse's potential contributions to psychology).

Conclusion: Individual multiverse projects are a meaningful alternative to traditional capstone projects, which do not require the collection of primary data and address many key undergraduate research methods learning outcomes.

Teaching Implications: We propose an evidence-based recipe for successful individual multiverse projects. In a post-COVID world, all project supervisors should have possibilities like this up their sleeves.

Keywords

Multiverse, dissertation project supervision, open science, statistics education

The choices we make when analyzing data shape the conclusions we draw. Collectively, these choices are referred to as *researcher degrees of freedom* (Simmons et al., 2011). At the analysis stage of a typical research project, there are many possible researcher degrees of freedom. For example, choices are made about which cases to exclude, how variables are wrangled, which inferential analyses are performed, and how Type 1 error control is handled. In a typical research project, just one set of choices is made, leading to just one set of conclusions. But, in most situations, a range of alternative choices can be made and justified[1]. In other words, in a typical research project, we take just one analytic path through a *garden of forking paths* (Gelman & Loken, 2013). However, there are many other possible and perfectly reasonable paths we could take.

In a multiverse analysis (Steegen et al., 2016), we make all combinations of all reasonable choices (within resourcing limits). In other words, rather than taking just one path through a garden of forking paths, we take all reasonable paths. Then, when the results are combined and synthesized, a sense of the robustness of a finding to analytic variability can be developed. Until recently, multiverse analyses have been the domain of experienced researchers, where the technique has been commonly applied to the reanalysis of previously published open data. Sometimes, these professional multiverses confirm the conclusions in the original publication (Godwin et al., 2025). Often, however, they do not (Credé & Phillips, 2017).

Multiverse analyses need not remain the domain of experienced researchers. Indeed, our experience in UK higher education suggests that multiverse studies can give students opportunities to practice a range of statistical techniques commonly taught in undergraduate psychology programs (TARG Meta-Research Group, 2022) and can help them meet many of the undergraduate psychology research methods learning outcomes specified by the British Psychological Society

[1]School of Psychological Science, University of Bristol, Bristol, UK
[2]Department of Psychology, James Cook University, Townsville, Australia

Corresponding Author:
Peter J. Allen, School of Psychological Science, University of Bristol, 12a Priory Rd, Bristol, BS8 1TU, UK.
Email: p.allen@bristol.ac.uk

(BPS, 2024). Finally, like computational reproducibility (Almuhanna & Allen, 2026) and other types of secondary-data projects, multiverse projects can provide a meaningful "equivalent alternative" (BPS, 2024, p. 19) to more traditional undergraduate capstone projects, which commonly involve face-to-face contact with human participants. Like traditional projects, secondary-data projects require students to critically engage with relevant literature, formulate research questions grounded in both theory and prior empirical evidence, justify methodological decisions, implement appropriate analyses, interpret results thoughtfully, and communicate their findings clearly and rigorously. In a post-COVID world, it is prudent to have such alternatives at hand.

This line of thinking led Heyman and Vanpaemel (2022) to develop *multiverse-in-the-classroom*. In their context (the Netherlands), psychology undergraduates undertake capstone projects that commonly comprise no more than one quarter of the third-year study load. Following some specialized training (on, e.g., open science and R Markdown), Heyman and Vanpaemel gave one group of eight students an open dataset and supported them through the process of designing, running, and reporting eight parallel multiverse projects. These multiverses contained between 18 and 160 outcomes. In some, all or most outcomes were statistically significant. In others, the *p*-values were much more variable.

According to Heyman and Vanpaemel (2022), this *many-multiverses-one-dataset* model affords a range of pedagogic benefits. These include opportunities to develop critical evaluation and data handling skills that reflect scientific best practice. However, because Heyman and Vanpaemel did not conduct a formal evaluation, the extent to which their students would have endorsed these benefits and/or demonstrated them in their project reports remains unknown. Additionally, although Heyman and Vanpaemel prefer the many-multiverses-one-dataset approach over one in which students individually choose a study to focus on, this approach denies students the opportunity to apply multiverse techniques to an effect of personal interest. It is known that topic choice, autonomy, and ownership are particularly valued by undergraduate project students (Todd et al., 2004). Finally, in the UK, the capstone project is typically larger than in the Netherlands and commonly comprises one-third of the final year study load. Furthermore, the BPS (2024) accreditation standards require that project students "individually… demonstrate a range of research skills including designing and planning, considering and resolving ethical issues, analysis and dissemination of findings" (p. 14). Within this context, the many-multiverses-one-dataset approach may not be substantial or individual enough.

However, Heyman and Vanpaemel (2022) briefly suggested an alternative for "students with a stronger background" (p. 3) that addresses these concerns. Specifically, they suggested that these students individually choose the effect on which to build a multiverse. For several years, PA has invited University of Bristol capstone project students to follow this suggestion. Specifically, these students have been charged with designing, running, and reporting individual multiverse projects in a supportive, small-group context. We invited one cohort of these students to reflect on their experiences qualitatively and used their reflections as the basis of an evaluation of student-led individual multiverse capstone projects. Furthermore, we combine our findings with several years of experience to propose an evidence-informed recipe for successful student multiverse projects in contexts like ours and outline the benefits and challenges of such projects for students and supervisors.

Method

Context

The BSc Psychology at the University of Bristol enrolls approximately 200 students annually and has a higher average Universities and Colleges Admissions Service (UCAS) tariff than most other UK undergraduate psychology courses (Complete University Guide Limited, 2026). The BSc is accredited by the BPS (2024), and one-third of each of the first two years of the program is dedicated to research methods training. One third of the final year is the capstone project, which is an "extensive piece of empirical research that requires the student to individually demonstrate a range of research skills including reviewing and establishing a theoretical understanding of the phenomenon of interest, formulating appropriate research questions or hypotheses, identifying an appropriate research design and associated methods of data collection, engaging in an ethical approval process, considering and resolving ethical issues and appropriately analysing and reporting the data in line with good scientific practice" (BPS, 2024, p. 19). Although capstone projects can be undertaken in groups, students are "required to demonstrate all of the above skills individually" (BPS, 2024, p. 19).

In the School of Psychological Science at the University of Bristol, academic staff supervise between five and seven capstone project students each year. Areas of expertise and proposed project outlines are advertised to students toward the end of their second year via short online videos. Students rank their top 10 preferences, and the Hungarian algorithm is used to maximize overall preference satisfaction while respecting supervisor capacity constraints. Typically, students will be assigned a supervisor within their top four preferences and meet with their supervisor and group once or twice before the end of their second year. In their third year, two hours of supervision per group are timetabled across each of the 20 teaching weeks. Although some supervisors divide the larger group into two or three smaller sub-groups and allocate supervision time accordingly, PA does not. However, within PA's weekly group supervision meetings, time is dedicated to both common themes and to the discussion of each student's individual project and progress. The principle underpinning this approach is that intellectual cross-fertilization benefits all group members. Additionally, the school occasionally offers optional "masterclasses" on topics likely to be of interest to the third-year cohort (e.g., ethics, data wrangling, and writing). The 6,000–7,000-word capstone project report is worth 40 of the 120 third-year credit points, representing around 400 h of work in total.

Table 1. Principles for Selecting an Effect for a Capstone Multiverse Project.

Principle	Rationale
Essential	
Personal interest	Given the time and intellectual investment required, developing a multiverse study around an effect of personal interest will help sustain motivation.
Open data that are sufficiently raw and complex	Developing a multiverse study around a published effect requires that the data underpinning the effect are open. Although it may be possible to source restricted (e.g., "available on request") data from the original researchers, this should not be relied on (Gabelica et al., 2022). Furthermore, it must be possible to imagine a range of sensible analytic alternatives using the data. Collecting original data on which a multiverse is based could guarantee sensible alternatives but may not be feasible for a capstone project.
Computationally reproducible effect	If the effect is not computationally reproducible it will not appear as an outcome within the garden of forking paths, and comparisons between it and a multiverse of alternative effects will not be possible. If a simple error (e.g., in reporting) is the probable cause of non-reproducibility, a decision about whether to use the corrected effect can be made.
Central effect(s)	It is easier to build a rationale for a multiverse study if the selected effect is central to the thesis of the paper in which it is nested. We suggest that the effect be reflected in the paper's title, abstract, and/or hypotheses, as well as reported prominently in the results section.
Desirable	
Single effect	Although a multiverse study can be developed around a small cluster of related effects, this can quickly add complexity to the project. In these situations, a smaller garden of forking paths is recommended.
Simple, familiar statistics	Students grapple with many novel concepts in a multiverse project. It is advantageous if these are applied in an already familiar, well-documented context. We've found that *t*-tests and correlation coefficients work well, although our students have also successfully worked with interaction terms and pairwise comparisons nested in ANOVAs, as well as regression coefficients.
Relatively large sample size	With a large sample, sub-sampling and exclusion criteria can be applied without sacrificing excessive statistical power. Such analytic choices may not be viable when the original sample size is small.
"Controversial" effect	Building a rationale for a multiverse study can be easier if there are doubts about the veracity of the effect. Doubts may arise due to a *p*-value close to .05, contradictory or inconclusive prior research, an unexpected finding, low statistical power, or suspicion about questionable research practices. However, other types of rationale are possible. For example, to understand how the magnitude of an effect that appears both solid and useful is influenced by different analytic choices.
Open code and/or codebook available	Reproducing the original effect(s), understanding how variables have been operationalized, and thinking about analytic alternatives are easier when the data and original analyses are clearly documented.

Of these, approximately 340–360 h are independent of the supervisor and other academic staff. Supervisors can discuss writing with students, but do not read drafts of students' project reports.

Collaborators

For this study, we collaborated with five former BSc Psychology students (EA, AB, RH, SM, and JW), who voluntarily contributed shortly after submitting their final project reports. All five students' second-year research methods unit marks were in the top quartile for their cohort, suggesting an aptitude for and interest in research methods and statistics. PA supervised them. Supervision meetings were weekly (for 20 weeks), group-based, and typically lasted two hours. Each student selected a published effect and built their multiverse project around it. To make these selections, they followed the collaboratively developed set of principles outlined in Table 1. The method they used to design their individual multiverse projects is described in Table 2, which was collaboratively developed during several supervision meetings. It should be noted that the actual design process was more iterative and recursive than can be illustrated in a table. For example, some development of the garden of forking paths typically occurred alongside the process of choosing an effect.

The students' final reports can be found in Hill et al. (2026). The findings they built their multiverse studies around included a lower willingness to engage with women wearing sexually suggestive t-shirts relative to athletic, academic, or plain t-shirts (Gurung et al., 2019); lower life satisfaction among participants who stopped using Facebook for five days compared with those who maintained normal usage (Vanman et al., 2018); a positive association between personal relative deprivation and self-reported aggression over time (Greitemeyer & Sagioglou, 2019); greater shifts toward counter-attitudinal positions on universal basic income among participants in control and perspective-taking/same-ideology conditions relative to those in a perspective-taking/different-ideology condition (Catapano et al., 2019); and a positive association between contact comfort with infection-risky others and perceived interpersonal value (Tybur et al., 2020). The students' gardens of forking paths contained between 48 and 84 analytic outcomes produced by systematically varying analytic sample specifications (using attention checks, careless-response indices, subgroup analyses, and power-based sample restrictions), the operationalization of variables and

Table 2. Method for Designing a Multiverse Capstone Project.

Activity	Questions and Issues to Consider
1. Initial literature familiarization	• *What is a multiverse analysis and how is multiverse situated within the open science movement?* • *What could a multiverse project look like?* Published examples can be mined for inspiration and provide a starting point for conversations about reasonable expectations for students.
2. Identify several possible effects	• *How can several suitable possibilities be rapidly identified?* Many journals use open science badges (see https://www.cos.io/initiatives/badges). Scanning their contents pages can be a better strategy than keyword searching. • *How can journals likely to contain papers with undergraduate-friendly statistics be identified?* Anecdotally, mid-ranked journals (see https://scimagojr.com) often contain a higher proportion of research based on undergraduate-level statistics than top-ranked journals. Similarly, we've found that the data and statistics used in social psychology research tend to be more manageable for undergraduates than those used in cognitive or individual differences psychology research. • *Are the data open and sufficiently complex? Is it possible to identify the key variables within the data file? If these are computed variables, are the variables used to construct them also available?* Unfortunately, open data badges do not guarantee that data are available and complete (Towse et al., 2021). Rapidly assessing the suitability of data can prevent spending too long pursuing ideas that are ultimately unworkable. • *To what extent can alternative analytic choices be imagined? Have these been considered elsewhere in the paper?* If few novel choices can be imagined, move to the next possibility.
3. Narrow down to one effect that is computationally reproducible	• *To what extent does each possibility meet the principles in Table 1?* Consider assigning a weight to each principle, scoring each effect using these weights, and then ranking them. • *Is the focal effect actually a cluster of linked effects (e.g., an interaction term and two simple effects)?* Consider this complexity when designing the garden of forking paths. • *What is the rationale for a multiverse project on this specific effect?*
4. Produce a garden of forking paths	• *What is a reasonable number of outcomes for the multiverse?* The size of the multiverse will trade-off against the complexity of implementing the alternative analytic choices. The resources available to students (including prior research methods training, the project timeline, supervisor availability, other competing commitments, etc.) should also inform decision making. • *Are the alternative analytic choices reasonable? Is there a rationale for each?* It should be possible to imagine a competent researcher making these choices (and all combinations of these choices). • *Are the alternative analytic choices feasible?* Some choices may be desirable, but not implementable given the design decisions made by the original researchers. Others may be challenging due to the technical expertise required to implement them. Testing the feasibility of analytic choices without PARKing (Pre-registering After the Results are Known; Yamada, 2018) can be achieved with shuffled data.
5. Pre-register	• *Should the study be formally pre-registered?* Although all research should begin with clear plans, some students may feel wary about making these public. If so, an informal pre-registration, whereby students file their plans with their school or supervisor, can be considered. Either way, acknowledging deviations from the plan or pre-registration in the final report is encouraged (Nosek et al., 2018) and can demonstrate critical reflection and attention to detail. • *How much detail should be included in the pre-registration?* Pre-registering a reporting plan as well as the aims and method can reduce the likelihood of selective reporting and/or emphasis in the final report. • *Should code be pre-registered?* If so, it can be created using shuffled data.

contrasts, the inclusion or exclusion of covariates, the handling of missing data and outliers, and the hypothesis-testing framework employed (i.e., frequentist or Bayesian). Across the five project reports, three students judged their focal effects to be largely robust across the analytic alternatives they considered (albeit with some variability in effect sizes and evidential strength), one concluded that the effect was not robust to analytic variation, and one characterized the evidence produced with their multiverse analysis as highly variable and ultimately non-diagnostic.

Procedure and Analysis

We used a qualitative inquiry methodology involving individual semi-structured interviews followed by one focus group, conducted using a conversational style to permit detailed exploration of the student-collaborators' lived experiences. The students were emailed participant information sheets prior to the interviews and focus group, and gave verbal consent at the start of each conversation. The James Cook

University Human Research Ethics Committee approved the research (ID: H8470).

AK had not met the students prior to this study and managed all recruitment, data collection, and analysis. Using Zoom, she undertook individual interviews with all five students ($M_{\text{length}} =$ 56 min) and then convened a focus group with three of them (64 min). The interview schedule (Allen et al., 2026) contained open questions designed to consider why the students were attracted to multiverse projects, the skills they developed, advice they would give to future students and supervisors, and the benefits and challenges of conducting an undergraduate multiverse project. Follow-up questions were asked to probe both positive and negative perspectives and experiences. In the focus group (Allen et al., 2026), the students discussed preliminary themes developed from the interviews.

The interviews and focus group were audio-recorded, transcribed verbatim, and deidentified, and the deidentified transcripts (Allen et al., 2026) were given to the students for member checking. Following Braun and Clarke's (2019) procedure for reflexive thematic analysis, AK familiarized herself with the data and then generated codes to identify themes. Using Microsoft Excel, AK used a reflexive and recursive approach to consider the data without a particular theoretical framework in mind. Initial codes drew on semantically similar content as well as exploration of implicit concepts. Related codes were clustered to develop broader themes and subthemes. After establishing the themes, AK revisited the data in full to consider diverse cases and minor themes relative to the developed themes and study aim. Data extracts, drawn from the whole corpus and selected to illustrate each finalized theme, are presented anonymously and reproduced "as spoken" with light editing and bracketed words added to improve readability. Acknowledging our implicit knowledge brought into the study (Braun & Clarke, 2019), we note that AK holds a PhD in psychology, works as an academic, and conducts research on the scholarship of teaching and learning. PA shares these characteristics, although he was not involved in data collection or analysis.

Results

The students were enthusiastic about conducting a multiverse capstone project. Their experiences and reflections are captured in two top-level themes: (a) factors that promote successful projects, and (b) the outcomes of these projects.

Factors Promoting Success

Intrapersonal, interpersonal, and contextual factors all contribute to successful multiverse projects. The students recognized that a multiverse project is "stats-oriented." Yet limited prior knowledge and confidence about statistics generally and the multiverse approach specifically were not seen as barriers. Acknowledging that students will find the idea and "language" of multiverse analyses novel, they recommended doing "a bit of reading on multiverse" as well as "familiarizing yourself with research methods and statistics a little bit more." Moreover, while the students suggested that having an interest in statistics might help, they all agreed that "anyone can probably have done this project with… support." They stated that this was because supervisors can work with students to "tailor the pathways of a multiverse to the level… and capability of the student."

While the students did not initially know each other, they embraced group supervision and met regularly outside of supervision. These meetings, in which they prepared and debriefed, meant that they could "get the best out of the [weekly supervision] meeting[s]." While each had their own topic and project, they credited regular group contact for promoting engagement, feeling "comfortable… to ask questions", and "keep[ing] each other on track." Regularly meeting also led to close friendships, within which students felt psychologically safe and emotionally supported ("you could bring up anything you were thinking about, anything you were struggling with, and we'd all kind of try and chip in and help each other"). They emphasized the value of group supervision while "doing separate studies", as there was a "benefit of working together and learning about this process [of designing and running a multiverse study] together." Supervisor support complemented peer support. As could be argued for all projects, a supervisor needs to "navigate that balance" between giving directed support and promoting autonomous discovery. This involves giving students "freedom" ("to explore ideas and to discuss [them] together"), while also being consistently present ("really being there and helping… not babying") and encouraging. The supervisor, as an "active, engaged partner" in the process, must notice when someone is struggling or not engaging. Additionally, for multiverse projects specifically, they'll also need "a good level of statistical knowledge", a "good grounding on what a multiverse is" and will need to practice "good science and transparency." The students believed these traits must be complemented by an awareness that students will need "guiding" in "something they've never done before" and, consequently, an ability to normalize making mistakes and "learn[ing] from them along the way."

Contextual factors promoting success included a solid plan. This was focused on devising a timeline to assist with meeting deadlines and milestones as with any capstone project (e.g., needing to be "quite organized and strategic, so that even if you don't have that knowledge, you're giving yourself the time…"), as well as a well-planned method for conducting the multiverse study. The students emphasized that the process of selecting an effect is not always "a straight path" and that "you can hit a lot of dead ends" without clear parameters. For example, "maybe you might use a paper and then two weeks down the line realize that you can't actually go ahead with it because there's an issue with the code or a decision point." As a result, Tables 1 and 2 offer "advice for people moving forward" by outlining the effect selection criteria and multiverse design process collaboratively developed and then followed by the students. Most notably, open data are essential. However, open data alone do not guarantee computational reproducibility, or that there will be "enough material… to understand and comprehend the data." Beyond the data, the size of the multiverse ("we had a limit on the amount of pathways that we were going to create") and statistical techniques involved ("sticking with slightly more simple forms of analysis"), the students stressed that personally interesting effects should be prioritized. In their words, being "really strict" about the criteria and process outlined in Tables 1 and 2 will help a multiverse project progress smoothly.

Multiverse Project Outcomes

The outcomes of multiverse capstone projects include student development (including skills and a sense of accomplishment) and the perceived contribution these projects can make to the discipline. Importantly, beyond the skills developed in any capstone project (e.g., writing and project management), students developed "a greater understanding of research methods" resulting in increased statistical literacy and confidence. Although a deeper understanding of a particular analysis can be expected from any capstone project, these students' statistical literacy and confidence transcended a single inferential analysis. This pertained to using specialized software as well as a deeper understanding of "researcher degrees of freedom", data analysis, and interpretation ("you start seeing all these relations between these different types of analysis… all of a sudden, these things that were very separate books almost become like a chapter in the same book"). They became more "critical" as researchers, and "look[ed] at research differently", having learned to question and "further critique" the nuance and impact of complex analytic choices. For example:

> I think the main skill was how I looked at the data, how you look at a research process in general. It's not just one path now for me. In my head, I think, "well what could I do differently? How many different paths could I be taking? How would this change the outcome?"

Moreover, their critical understanding of the impact of analytic choices and reporting practices underpinned perceived points of difference from peers who had not done multiverse projects (where "the statistics [are] a means to an end" for discovering whether the "hypothesis is or isn't significant").

Indeed, when asked to compare multiverse to traditional capstone projects, the students recognized far more strengths than downsides. They pointed to feeling an "extra level of security" in completing their projects because "you're using data that's already been sourced" and because you are "the main driving force." In other words, external challenges like participant recruitment difficulties or bugs in digital data collection tools are not a concern, because the project is "within your circle of control." When reflecting on multiverse relative to traditional capstone projects, the students did not feel they were disadvantaged or "missing out on anything by choosing to do a multiverse." However, they acknowledged that multiverse projects using open data did not involve designing and conducting a new primary data collection study. For some students, this was not a concern given earlier experiences in their degree; yet one student expressed that, upon graduating, they had "not yet done" data collection outside of first- and second-year research methods class projects. They appreciated that some students may want to "build an experiment from the ground up", and that while their projects did not have a data collection component, a multiverse project could conceivably be designed to "collect your own data as a group and work on one dataset."

The students also described developing a "skill set" that stretched beyond single skills, framed as demonstrating core graduate competencies, or the "requirements for what's expected of a psychology undergraduate." In their words, "multiverse [projects are] really good at teaching a lot of the skills which are expected" as a professional in the discipline. For example, one student argued that they had met "five of the seven" standards for transparent research proposed by Wagenmakers et al. (2021). In this way, the students understood their project not only as an assessment but as a vehicle for the development of the professional skills it is designed to develop. Moreover, their application of the skills learned transcended the completion of their project:

> Now I'm still putting into practice in life when it comes to reading articles or even headlines or things like that – it's just being a lot more understanding and critical of the information that you're being given.

While completing their projects was "rewarding" and led to feelings of "accomplishment" and "achievement", as could be expected from completing any capstone project, the students also perceived multiverse projects as "really interesting and very, very enjoyable." This positivity was long-lasting and was considered as a point of difference from their peers' experiences. As one remarked,

> I learned a huge amount… and I know a lot of people who've done their [capstone] dissertation probably don't feel quite the same way…. I know a lot of them don't even want to think about what they've done anymore… [but] I don't feel drained at all around the topic; it's still something that I'm really interested in and in all of the things that I've learned from it.

They felt pride from their "perseverance" and from successfully running analyses and interpreting and communicating their findings. Their increased confidence in "a whole set of skills that you have as a researcher" also resulted in developing an increased sense of professional identity. For some, a new facet to their identity involved a passion for advocating better discipline practices and the role that they can play (multiverse "can help change the sort of problem[s] within science… we can improve the discipline").

Indeed, all the students recognized how multiverse projects can benefit the discipline. First, because a multiverse illustrates how statistical choices are integral to psychological knowledge production, "it's exposing to students (and then hopefully future generations of researchers and even researchers in the field now) [how] certain analytic decisions can have a really big impact on an effect." Additionally, misunderstandings about power and misconceptions about *p*-values can be addressed: "you can look at the different ways [that different power levels] affect the effect…. You can compare Bayesian and frequentist [models] and… looking at them together actually makes you understand them more." Thus, multiverse projects can both "capture the subjectivity" reflected in researcher degrees of freedom and help move the discipline beyond "binary thinking" about statistical significance.

More broadly, and in light of the replication crisis, the students saw multiverse as a "tool" that can signal and promote credibility: "it's… a stamp on your main effect, isn't it, to say that 'within that dataset I can be confident that this is what was going on.'" There was further recognition of how this could "help get rid of this biased literature" and "file drawer" effects. To that end, they felt that multiverse alongside other transparent decision-making initiatives (e.g., pre-registration) could "help to calibrate… the scientific ecosystem to higher standards of efficiency, quality, and credibility." It is important to note that, even when students encountered researchers who did not reply to analysis queries or discovered errors in open data files and published papers (e.g., "even when data is open, it's not necessarily as open as it may seem"), they did not feel discouraged by their discipline. Rather they learned to prioritize and value transparency and credibility. Through the multiverse process, they learned that "methods [are] such a big part of psychology," which contextualized larger discipline debates:

> You can become nihilistic about research sometimes when you just get told that nothing replicates and psychology is a failure…. But I think when you deal with these kinds of issues, you realize it's a lot more nuanced. And that there are solutions.

Thus, rather than being defeated by results that did not reproduce, the students focused on how open data, pre-registration, and transparent analyses (inclusive of multiverse analyses) could improve the reputation of the discipline and the trustworthiness of its findings: "to help the research that we are now producing be as valid and reliable as it possibly can be."

Discussion

Prior to the current study, Heyman and Vanpaemel (2022) provided the only known discussion of multiverse in the context of undergraduate psychology education. They described several pedagogic benefits of their many-multiverses-one-dataset approach to undergraduate capstone multiverse projects but did not formally evaluate it. We have adapted their model to accommodate a higher level of individualization and qualitatively evaluated it with one full cohort of project students. This evaluation indicated that factors promoting success in the students' individual projects were intrapersonal, interpersonal, and contextual. It is, however, important to note how the students stressed that it "wasn't one thing that made this a good experience," rather it was "everything in combination." Nonetheless, they emphasized the important elements were in their control. Importantly, they endorsed the idea that "anyone could do a multiverse" with appropriate scaffolding. The students also spoke to many skills that can be developed through such projects (e.g., statistical literacy and confidence; critical appraisal of research evidence and claims) and the potential contribution that multiverse projects can make to psychology (inclusive of tackling big discipline issues around transparency and reproducibility). Such potential can be further extended by working with motivated students to convert their project reports into peer-reviewed publications (e.g., Fowles & Allen, 2026).

Based on our findings and experience, we propose that project supervisors seeking to further adapt and test the idea of multiverse capstone projects begin by giving students opportunities to develop as a team whilst collaboratively completing activity 1 in Table 2. The students could then be tasked with producing a list of criteria like that in Table 1, and a method like that described in the remainder of Table 2. Once the plan is clear, they can begin executing it. At this stage, students should be reminded that the path from an initial idea to a finished research report is rarely linear. It is likely that they will encounter some dead ends and false starts (e.g., even experts find computational reproducibility challenging; see, for example, Hardwicke et al., 2021), and that their actual approach is likely to be iterative and recursive. This is common to all research projects. Through engaging with this process, we anticipate that students will derive a range of benefits. For example, they will have opportunities to practice both subject specific (e.g., a range of statistical techniques) and transferrable (e.g., project management) skills aligned with the graduate attributes commonly specified by accreditors and awarding bodies, and widely valued by employers (Serbic & Bourne, 2020). They will develop a nuanced understanding of best (open and transparent) practices in psychological research, and a critical awareness of how researcher choices influence research outcomes. More broadly, individual multiverse projects allow students to pursue personal interests within a shared framework, are inclusive and "COVID-safe," and repurpose existing data, which can take pressure off subject pools and relieve students from one of the most common stress points in a traditional undergraduate capstone project – recruiting a large enough sample.

Realistically, a substantial part of a full academic year is needed for projects like those described herein. If working to a shorter timescale, the process could begin by giving students this paper and drawing their attention to Tables 1 and 2. Such an approach, perhaps combined with Heyman and Vanpaemel's (2022) many-multiverses-one-dataset model, may also be necessary for less experienced, able or motivated students, or in schools where less academic time is allocated to supervision. Alternatively, students could be given the choice between several effects that have been pre-selected to ensure computational reproducibility, a range of viable analytic alternatives, and a level of challenge commensurate with the students' experience and the available supervision time. To reduce workloads further and/or to keep the cognitive overhead manageable for supervisors of larger groups of capstone students, the students could choose papers/effects in pairs/small groups. Conversely, for stronger students, where a more substantial project is desired, and/or a more intensive level of supervision is possible, crowdsourcing analytic alternatives for the multiverse (see Heyman et al., 2025) could provide an interesting challenge and some data collection experience, the latter of which our students felt was lacking in their own projects. Alternatively, a modest traditional study (e.g., an online survey study) could be combined with the many-multiverses-one-dataset model if timelines permit and a more extensive data collection experience is desired. We have not yet had opportunities to put these ideas into practice.

Further thought should also be given to the feasibility of multiverse projects when students are allocated to, rather than self-select, projects and supervisors. Our student-collaborators felt that a successful multiverse project, when appropriately scaffolded, would be interesting and achievable for a wide range of students. However, this is yet to be tested. Finally, educators considering multiverse capstone projects should first assess the extent to which such projects are aligned with their own institutional and accreditation expectations. We feel confident that both our approach and that described by Heyman and Vanpaemel (2022) can be adapted to meet a range of different demands and circumstances. However, this requires further consideration.

We acknowledge that the qualitative approach we used involved gathering perspectives from students who undertook a multiverse capstone project, and that we do not have comparison data. It falls on future researchers to directly compare the skills and experiences of students completing multiverse projects to more traditional capstone projects. Such work could also explicitly consider different types of multiverse project configurations as well as their viability when supervisors and students have varying levels of interest and experience with statistics and these kinds of projects. Additional methodologies could also triangulate findings. For example, anonymous surveys with an entire cohort and/or repeated measures designs to explore anticipated and realized challenges and outcomes would advance our understanding of how best to design future multiverse capstone projects.

In conclusion, this paper proposes and demonstrates that individual multiverse projects can be a meaningful alternative to more traditional capstone projects and can address many key undergraduate research methods learning outcomes. In a post-COVID world, all project supervisors should have possibilities like this up their sleeves.

ORCID iDs

Peter J. Allen https://orcid.org/0000-0002-9690-1545
Emily Abraham https://orcid.org/0009-0007-0989-7271
Rory Hill https://orcid.org/0009-0002-8898-1714
Amanda E. Krause https://orcid.org/0000-0003-3049-9220

Author Contributions

PA contributed to conceptualization, methodology, writing–original draft, writing–review and editing, project administration. EA, AB, RH, SM, and JW contributed to conceptualization, writing–review and editing. AK contributed to conceptualization, methodology, investigation, data curation, formal analysis, writing–original draft, writing–review and editing. EA, AB, RH, SM, and JW contributed equally to this manuscript and are ordered alphabetically.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Transparency and Openness Statement

The interview schedule, focus group schedule, and raw data underpinning this research are available at https://osf.io/fjw42/.

Note

1. This point is illustrated vividly in *many analysts* projects like Hoogeveen et al. (2023) and Aczel et al. (2026).

References

Aczel, B., Szaszi, B., Clelland, H. T., Kovacs, M., Schulz-Kümpel, H., Holzmeister, F., Nilsonne, G., Hoffmann, S., Kosa, L., Torma, Z. A., Abdelfatah, Y., Aberson, C. N., Acar, O. A., Acem, E., Adamkovic, M., Adamovich, T., Adiasto, K., Ahnström, L., Akil, A. M., & Nosek, B. A. (2026). Investigating the analytical robustness of the social and behavioural sciences. *Nature*, *652*, 135–142. https://doi.org/10.1038/s41586-025-09844-9

Allen, P. J., Abraham, E., Bleach, A., Hill, R., Mahmud, S., Wild, J., & Krause, A. E. (2026). "Anyone could do this!": Multiverse as an undergraduate psychology capstone project (online supplement). https://osf.io/fjw42/

Almuhanna, A., & Allen, P. J. (2026). The computational reproducibility of psychology learning and teaching research. *Teaching of Psychology*. Advance online publication. https://doi.org/10.1177/00986283251412796

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, *11*(4), 589–597. https://doi.org/10.1080/2159676X.2019.1628806

British Psychological Society. (2024). Standards for the accreditation of undergraduate, conversion and integrated Master's programmes in psychology. https://www.bps.org.uk/accreditation/education-providers

Catapano, R., Tormala, Z. L., & Rucker, D. D. (2019). Perspective taking and self-persuasion: Why "putting yourself in their shoes" reduces openness to attitude change. *Psychological Science*, *30*(3), 424–435. https://doi.org/10.1177/0956797618822697

Complete University Guide Limited. (2026). Psychology: Subject league table 2026. https://www.thecompleteuniversityguide.co.uk/league-tables/rankings/psychology

Credé, M., & Phillips, L. A. (2017). Revisiting the power pose effect: How robust are the results reported by Carney, Cuddy, and Yap (2010) to data analytic decisions? *Social Psychological and*

Personality Science, *8*(5), 493–499. https://doi.org/10.1177/1948550617714584

Fowles, H. E., & Allen, P. J. (2026). Multiverse analysis of the effect of a non-deceptive placebo on a neural measure of emotional distress. *Journal of Robustness Reports*, Article 3-rr1. https://doi.org/10.21468/JRobustRep.3-rr1

Gabelica, M., Bojčić, R., & Puljak, L. (2022). Many researchers were not compliant with their published data sharing statement: A mixed-methods study. *Journal of Clinical Epidemiology*, *150*, 33–41. https://doi.org/10.1016/j.jclinepi.2022.05.019

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

Godwin, H. J., Lee, C. E., & Drieghe, D. (2025). A multiverse analysis of cleaning and analyzing procedures of eye movement data during reading. *Behavior Research Methods*, *57*, Article 164. https://doi.org/10.3758/s13428-025-02689-0

Greitemeyer, T., & Sagioglou, C. (2019). The impact of personal relative deprivation on aggression over time. *The Journal of Social Psychology*, *159*(6), 664–675. https://doi.org/10.1080/00224545.2018.1549013

Gurung, R. A. R., Stoa, R., Nelson, A., & Schultz, D. (2019). The influence of t-shirt graphics on perceptions and behavioral intentions. *The Journal of Social Psychology*, *159*(3), 328–339. https://doi.org/10.1080/00224545.2018.1491471

Hardwicke, T. E., Bohn, M., MacDonald, K., Hembacher, E., Nuijten, M. B., Peloquin, B. N., deMayo, B. E., Long, B., Yoon, E. J., & Frank, M. C. (2021). Analytic reproducibility in articles receiving open data badges at the journal psychological science: An observational study. *Royal Society Open Science*, *8*(1), Article 201494. https://doi.org/10.1098/rsos.201494

Heyman, T., Pronizius, E., Lewis, S. C., Acar, O. A., Adamkovič, M., Ambrosini, E., Antfolk, J., Barzykowski, K., Baskin, E., Batres, C., Boucher, L., Boudesseul, J., Brandstätter, E., Collins, W. M., Filipović Ðurđević, D., Egan, C., Era, V., Ferreira, P., Fini, C., & Buchanan, E. M. (2025). Crowdsourcing multiverse analyses to explore the impact of different data-processing and analysis decisions: A tutorial. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000770

Heyman, T., & Vanpaemel, W. (2022). Multiverse analyses in the classroom. *Meta-Psychology*, *6*, Article MP.2020.2718. https://doi.org/10.15626/MP.2020.2718

Hill, R., Bleach, A., Mahmud, S., Wild, J., Abraham, E., & Allen, P. J. (2026). Multiverse analyses – A group undergraduate project. University of Bristol. https://osf.io/9kcgm/

Hoogeveen, S., Sarafoglou, A., Aczel, B., Aditya, Y., Alayan, A. J., Allen, P. J., Altay, S., Alzahawi, S., Amir, Y., Anthony, F.-V., Appiah, O. K., Atkinson, Q. D., Baimel, A., Balkaya-Ince, M., Balsamo, M., Banker, S., Bartoš, F., Becerra, M., Beffara, B., & Wagenmakers, E.-J. (2023). A many-analysts approach to the relation between religiosity and well-being. *Religion, Brain & Behavior*, *13*(3), 237–283. https://doi.org/10.1080/2153599X.2022.2070255

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(11), 2600–2606. https://doi.org/10.1073/pnas.1708274114

Serbic, D., & Bourne, V. (2020). Final year research project as a tool for maximising students' employability prospects. *Psychology Teaching Review*, *26*(1), 90–95. https://doi.org/10.53841/bpsptr.2020.26.1.90

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. https://doi.org/10.1177/1745691616658637

TARG Meta-Research Group. (2022). Statistics education in undergraduate psychology: A survey of UK curricula. *Collabra: Psychology*, *8*(1), Article 38037. https://doi.org/10.1525/collabra.38037

Todd, M., Bannister, P., & Clegg, S. (2004). Independent inquiry and the undergraduate dissertation: Perceptions and experiences of final-year social science students. *Assessment & Evaluation in Higher Education*, *29*(3), 335–355. https://doi.org/10.1080/0260293042000188285

Towse, J. N., Ellis, D. A., & Towse, A. S. (2021). Opening Pandora's Box: Peeking inside psychology's data-sharing practices, and seven recommendations for change. *Behavior Research Methods*, *53*(4), 1455–1468. https://doi.org/10.3758/s13428-020-01486-1

Tybur, J. M., Lieberman, D., Fan, L., Kupfer, T. R., & de Vries, R. E. (2020). Behavioral immune trade-offs: Interpersonal value relaxes social pathogen avoidance. *Psychological Science*, *31*(10), 1211–1221. https://doi.org/10.1177/0956797620960011

Vanman, E. J., Baker, R., & Tobin, S. J. (2018). The burden of online friends: The effects of giving up Facebook on stress and well-being. *The Journal of Social Psychology*, *158*(4), 496–508. https://doi.org/10.1080/00224545.2018.1453467

Wagenmakers, E.-J., Sarafoglou, A., Aarts, S., Albers, C., Algermissen, J., Bahník, Š, van Dongen, N., Hoekstra, R., Moreau, D., van Ravenzwaaij, D., Sluga, A., Stanke, F., Tendeiro, J., & Aczel, B. (2021). Seven steps toward more transparency in statistical practice. *Nature Human Behaviour*, *5*, 1473–1480. https://doi.org/10.1038/s41562-021-01211-8

Yamada, Y. (2018). How to crack pre-registration: Toward transparent and open science. *Frontiers in Psychology*, *9*, Article 1831. https://doi.org/10.3389/fpsyg.2018.01831