

Developing and Piloting a Quick Evaluation of Research Integrity for Evidence Synthesis in Education (QUERIES-Edu)

Authors:

Lucija Batinović^{1*}, Marta Topor¹, Natalie Hyltse², Henrik Danielsson¹, Nicholas JL Brown², Jamie Cummins^{3,4}, Adrien Fillon⁵, James Heathers², Gustav Nilsson^{6,7}, Samuel J Westwood¹¹, Laura Caquelin⁶, André Kalmendal², Thomas Nordström², Eli Thoré^{8,9,10}, Rickard Carlsson²

Affiliations:

¹Department of Behavioural Sciences and Learning, Linköping University

²Department of Psychology, Linnaeus University, Sweden

³Department of Psychology of Digitalisation, University of Bern, Switzerland

⁴Department for Consumer Behaviour, University of Bern, Switzerland

⁵LAPSCO–UMR UCA-CNRS 6024, Clermont-Ferrand University, France

⁶Department of Clinical Neuroscience, Karolinska Institutet, Sweden

⁷Department of Psychology, Stockholm University, Sweden

⁸Laboratory of Adaptive Biodynamics, Research Unit in Environmental and Evolutionary Biology, Institute of Life, Earth, and Environment, University of Namur, Belgium

⁹Department of Wildlife, Fish, and Environmental Studies, Swedish University of Agricultural Sciences, Sweden

¹⁰TRANfarm–Science, Engineering, & Technology Group, KU Leuven, Belgium

¹¹Department of Psychology, Institute of Psychiatry, Psychology, Neuroscience, King's College London, London, United Kingdom

*** Corresponding author:** lucija.batinovic@liu.se

Abstract

The threat to evidence synthesis is becoming more prominent with increasing worries that the unexpected amounts of retrieved studies, questionable data and inconsistent findings are beginning to overwhelm the publication system and potentially bias meta-analytic conclusions. This paper presents an integrity assessment checklist (QUick Evaluation of Research Integrity for Evidence Synthesis in Education; QUERIES-Edu) and a pilot study plan that provides an overview of research integrity status in special education research and assesses the feasibility of the checklist. The checklist consists of questions that evaluate the research integrity of a scientific study, as well as its journal and publisher. An expert panel will provide feedback and suggestions for the items, after which the final version will be implemented. The checklist will be evaluated by a user panel that will conduct the assessment on studies included in a large scoping review of educational interventions for students with intellectual disability. Our aim is that the checklist will enable authors of systematic reviews to conduct quick checks of studies before they are included in thorough quality appraisal, which in turn will facilitate evidence synthesis and ensure the trustworthiness of included studies.

Keywords: trustworthiness, credibility, research integrity, error detection

Introduction

The growing awareness of the replication crisis in Psychology and in science more generally is well into its second decade (Ioannidis, 2005; Simmons et al., 2011; Open Science Collaboration, 2015). Paired with an increased public recognition of “questionable research practices” (John et al., 2012), “paper mills” (Abalkina & Bishop, 2023), predatory publishing (Beall, 2012), and, more recently, fully AI-generated, fraudulent papers (Liverpool, 2023), the integrity and transparency of scientific research has been called into question.

Issues of untrustworthy research not only threaten the credibility of individual studies; they directly reduce the quality of *evidence syntheses*, such as systematic reviews and meta-analyses. Importantly, these evidence syntheses are considered the gold standard of evidence in many research areas (e.g., Medicine, Psychology), and they have a major influence on real-life policy and professional practice. When erroneous or fraudulent research seeps into systematic reviews and meta-analyses, the results will be, at best, less accurate, and at worst, misleading.

While the issue of problematic studies is likely ubiquitous throughout all of educational research, this project focuses on Special Education and Disability. Within these areas, there can undoubtedly be detrimental consequences if professional recommendations and policies are informed by unreliable or misleading evidence. Thus, the aim of this project is to develop and validate a checklist to identify and evaluate problematic studies (and journals) in Special Education and Disability research, as a way to retain as much research integrity as possible in evidence syntheses.

Background

Great efforts have been exerted into regaining trust in research findings (e.g., the [Center for Open Science](#); pre-registrations and registered reports, Chambers & Tzavella, 2021; guides on justifying sample sizes and alpha levels, Lakens et al., 2018), and the Open Science movement has highlighted the importance of data sharing and pre-registering hypotheses to mitigate biases and improve the quality of our work (Bertram et al., 2023). Unfortunately, not only have questionable research practices not yet been eliminated, but recently another type of issue has started to be discovered. The proliferation of AI tools has exacerbated the number of fake papers and paper mills (Abalkina & Bishop, 2023; Liverpool, 2023; Heathers, 2024).

Every single published study is a potential source of evidence that can be included in a systematic review or meta-analysis, the results of which, in turn, may dictate the policies of governments, healthcare systems, and other institutions (Lakens et al., 2016). If studies are not conducted and reported accurately and honestly, the accuracy and trustworthiness of evidence synthesis will be reduced (the so-called “garbage in—garbage out” problem; see e.g., Sotola, 2022). Furthermore, questionably conducted and fraudulent studies are likely faster to produce in comparison to legitimate research, as well as requiring less time and effort to publish (e.g., due to predatory journals and paper mills; e.g., Beall, 2012; Else & van Noorden, 2021). Indeed, there are growing worries that unexpectedly large numbers of studies retrieved by citation database searches, questionable data, and inconsistent findings

are beginning to overwhelm the publication system (Else, 2024). As such, studies with questionable research integrity not only introduce doubt into the existing evidence, but also lengthen and obstruct the process of conducting a systematic review, which is in itself already a time-consuming and difficult endeavor.

Rationale for the QUERIES-Edu Checklist

During a scoping review of school-based interventions for students with intellectual disabilities (Batinović et al., 2025), we identified multiple recurring issues while screening and coding the articles. This led us to shift from our original aim, which was to assess methodological quality, to evaluating research integrity and transparency. Preliminary findings indicated a low adherence to open science practices and inadequate reporting of ethical approval, prompting a systematic evaluation of other problematic aspects of this body of literature.

INSPECT-SR (INveStigating ProBlEmatic Clinical Trials in Systematic Reviews) is a tool recently developed to assess the integrity of randomized clinical trials (RCTs) in the context of systematic reviews (Wilkinson et al., 2024). While the INSPECT-SR project is a major inspiration for the present checklist, the tool is specifically designed for RCTs; many of its items are inapplicable for educational interventions in disability research. Therefore, we decided to create another checklist that is relevant and applicable to educational research.

By compiling and cross-referencing items from existing checklists and previous reviews, we produced a new combined list of relevant items. Along with new items deemed relevant by the team that conducted the scoping review, a preliminary list of items that estimate a study's research integrity was assembled. Additionally, we believe that assessing the publishing venue (i.e., journals and their publishers) is equally important as examining the paper itself. Thus, we paid extra attention to identifying previous items and creating new items that evaluate the trustworthiness of both the paper and the journal it was published in.

Existing Tools and Gaps

As mentioned above, the initial, preliminary QUERIES-Edu checklist is largely developed by compiling and cross-referencing items from existing tools in combination with new, relevant items. Thus, overlap can be seen between previous tools and QUERIES-Edu. There are, however, clear gaps in the existing work that QUERIES-Edu will be able to fill.

While INSPECT-SR (Wilkinson et al., 2024) is likely the most large-scale, comprehensive, and presently perhaps most influential checklist, other tools for evaluating the trustworthiness of published science have been developed in the past decade, albeit more modest in scope. Some of the more prominent ones are the TRACT checklist (Mol et al., 2023) and REAPPRAISED (Grey et al., 2020), which are aimed at biomedical RCTs, and QuOCCA (Héroux et al., 2022) which is aimed at overall research quality and reproducibility in biomedical publications. Some of these tools are part of the foundation of INSPECT-SR, which consists of an amalgamation of much of previous work on quality and trustworthiness assessments within the medical field.

The most notable gap for QUERIES-Edu to fill is the intended research field of application: all existing tools mentioned above have a focus on medicine. Additionally, given their medical focus, most tools primarily apply to randomized controlled trials. As such, QUERIES-Edu can provide a long-overdue alternative way to assess research integrity that is specifically developed to be relevant to social sciences, and education in particular, as well as for non-medical interventional studies.

In regards to journal-level assessments, tools include the Journal Evaluation Tool by Rele et al. (2017) and a compilation of frameworks and checklists by Frandsen (2019).

Aim and Objectives

We will develop the QUick Evaluation of Research Integrity in For Evidence Synthesis in Education (QUERIES-Edu) checklist with systematic reviews and meta-analyses as our target area of use, and authors of systematic reviews as primary users. The aim of the checklist is to identify studies, including non-randomized controlled studies, that should *not* be included in an evidence synthesis, as the lack of accuracy, honesty, and/or trustworthiness of the study's results can only contaminate the synthesis. QUERIES-Edu is thus developed to be conducted *before* the traditional quality and methodology assessment which applies conventional tools such as Risk of Bias 2 (Sterne et al., 2019) and CEC Quality Indicators (Gersten et al., 2005). As such, this checklist could be considered as a part of the screening stage, where studies are still excluded, rather than the quality assessment stage, where studies are graded/scored (on items such as pre-registration, clearly stated inclusion and exclusion criteria, and exhaustive searches), but traditionally are included independent of the assessment grade/score.

QUERIES-Edu will focus on a smaller subset of items drawn from various sources, including our own suggestions and multiple panel evaluations, that allow a brief, superficial estimation of trustworthiness of the study. Therefore, it is essential that the checklist be easy to use and quick at detecting problematic studies.

The primary scope for this checklist is educational interventions, and our main concern for the items will be whether they cover all and any problematic aspects in studies in this field. We expect a large overlap of content for interventions done in any area of education research, thus we will pilot the checklist on a sample of interventions from special education research.

Following this premise, our objectives are:

1. Develop a checklist that can detect problematic intervention studies in education research.
2. Develop a checklist that can detect journals and publishers with problematic business practices.

Methodology

Supplemental materials are available on [OSF](#).

Recruitment of Panels

We have recruited a panel of experts, who will serve as reviewers providing additional feedback and suggestions for improving the checklist, see Table 1 for inclusion criteria. We recruited six panelists for the expert panel.

Additionally, we are recruiting a panel of end users, who will pilot the use of the checklist on various studies. Our aim for the User Panel is to recruit a diverse set of raters, including researchers at various career levels and with various expertise of systematic reviews and the research domain, see Table 1 for inclusion criteria. A diverse panel will ensure that the researcher group piloting of the QUERIES-Edu checklist is as representative of the intended end users as possible, that the checklist is understood properly with different expertise levels, and that it can be successfully applied without extensive experience in assessment of research integrity. So far, we have recruited four researchers to the User Panel, but the recruitment will continue during Phase 2 as well (see Figure 1).

All panelists were recruited through personal networks of professionals and chain-referrals. Some panelists were recruited from the list of interested collaborators from an earlier hackathon (SIPS, 2023 in Padua, Italy) related to the development of a similar checklist.

Table 1.
Inclusion criteria for expert and user panels

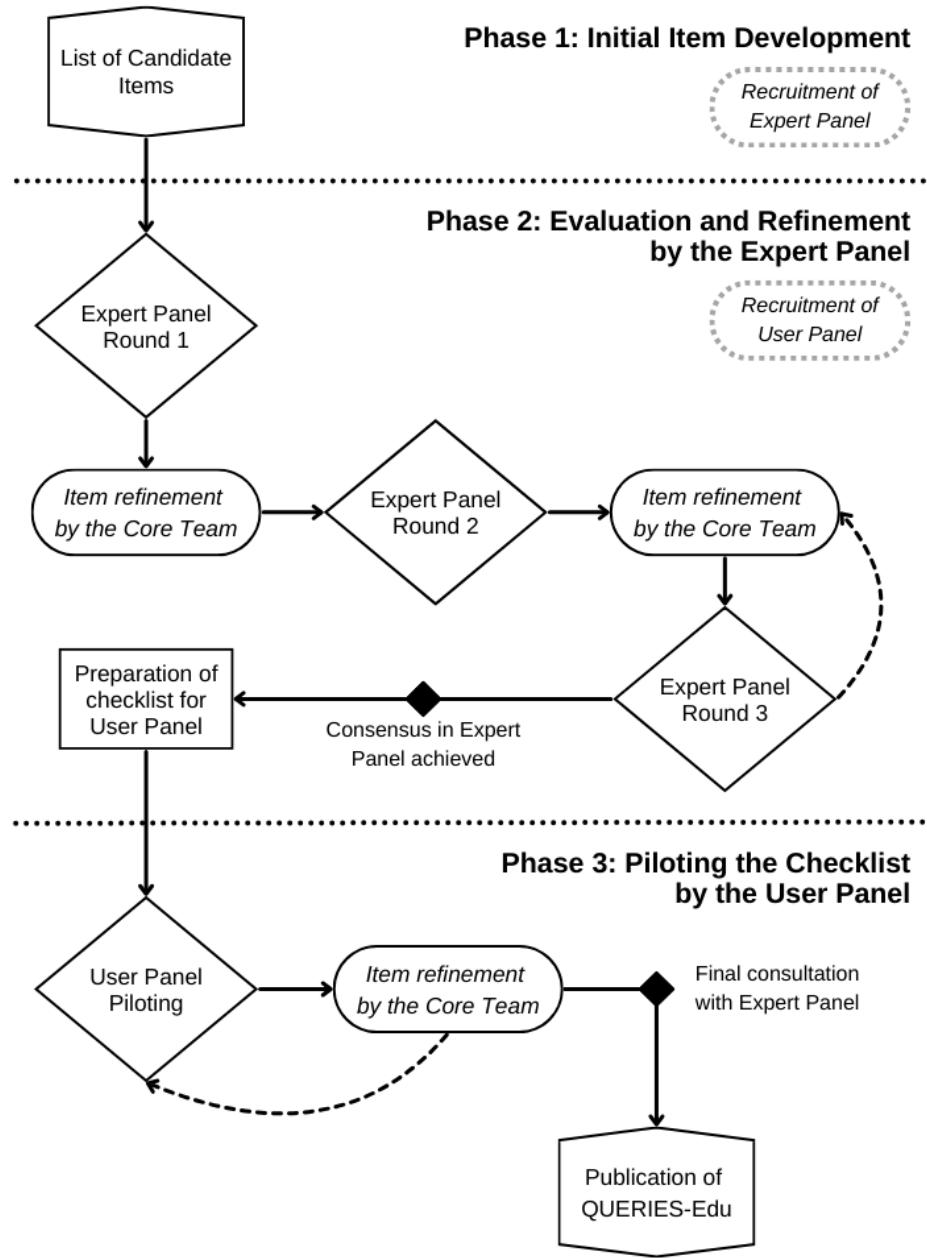
	Inclusion Criteria
Expert Panel <i>Recruited during Phase 1</i>	Expertise in research integrity evaluation ● Involvement in developing or applying tools for evaluating research integrity and/or similar (e.g., error detection) <i>and/or</i> Domain expertise ● Domain-specific knowledge of educational and/or psychological interventions.
User Panel <i>Will be recruited during Phase 2</i>	Career stages and levels of expertise Aim: diversity among raters/users ● Various career levels (from early career to senior researcher) ● Various expertise (both in conducting systematic reviews and meta-analyses, and domain-specific research)

Methods for Developing the Checklist

As illustrated in Figure 1, the checklist development process consists of three phases: Phase 1: Initial Item Development, Phase 2: Evaluation and Refinement of the Checklist by the Expert Panel, and Phase 3: Piloting the Checklist by the User Panel. We are currently at the

end of Phase 1, soon ready to distribute the initial list of candidate items to the Expert Panel and launch Phase 2.

Figure 1
Flow diagram of the development process.



Phase 1: Initial Item Development

Two authors (LB and MT) conducted the first compilation of potentially relevant items during the data extraction process of a scoping review (Batinović et al., 2025), with guidance from research integrity experts. They aggregated items from tools mentioned in the [Existing Tools and Gaps](#) section and items that were flagged as important during the data extraction

process. The list initially had 180 items, from which they extracted the most relevant items that went into the checklist (available on [OSF](#)).

Item development was guided by the aim of detecting “problematic” studies. We followed the INSPECT-SR framework and defined “problematic” studies as *‘any published or unpublished study where there are serious questions about the trustworthiness of the data or findings, regardless of whether the study has been formally retracted’* based on Cochrane Policy on identifying problematic studies (Boughton et al, 2021; Cochrane, 2021).

This was then further refined by two co-authors (RC and NH). While they had experience from the INSPECT-SR project (Wilkinson et al., 2024) and the initial development of another, similar checklist, they were new to this specific list of items. Together with another co-author (GN), they hosted an unconference workshop at the SIPS conference in Padua, 2023 on the development of an error detection checklist (Hyltse et al., 2023). When refining the present checklist, they were able to incorporate feedback from the workshop participants on various items (primarily from the first candidate list of items from the INSPECT-SR project). This resulted in a checklist of 36 candidate items (see Appendix 1. List of Candidate Checklist Items) that will be presented to the Expert Panel (recruited during Phase 1) during the first round of review, at the beginning of Phase 2 (see Figure 1).

Checklist Structure

At the end of Phase 1, we have created a candidate list of items that will eventually go into a brief checklist for evaluating scientific integrity, mainly designed for assessments of primary studies and for determining whether they should be included in evidence synthesis or not. We intend the QUERIES-Edu checklist to evaluate two domains: Article Assessment and Journal/Publisher Assessment, with items grouped into subdomains. The finalized checklist (distributed to the User Panel in Phase 3, see Figure 1) will include a guidance document containing the items, elaboration of each item, and the scoring system. Item descriptions, scoring, and guidance will be developed by the core team, incorporating feedback from the Expert Panel. While the iterative review process with the Expert and User panels will dictate the inclusion of items, and in turn the grouped subdomains as well, see Table 2 for the two domains and preliminary subdomains of interest. See also list of potential items (Appendix 1) and the panel sheets (available on [OSF](#)).

Table 2.

Proposed domains and subdomains of interest

Domain	Subdomains
Article assessment	● Integrity validity ● Plausibility ● Plagiarism ● Consistency ● Transparency ● Additional checks
Journal/publisher assessment	● Business model of the Publisher ● Communication of the journal ● Business model of the journal ● Transparency of the journal ● Integrity of the journal ● Additional checks ● Editorial process of the journal

Phase 1 has been finalized prior to Stage 1 RR submission, and the following phases will be conducted after receiving in-principle acceptance.

Phase 2: Evaluation and Refinement of the Checklist by the Expert Panel

During Phase 2 (see Figure 1), the Expert Panel will evaluate and refine the proposed checklist. As the checklist has already gone through extensive work, we anticipate that the panel will view it as quite exhaustive. However, we also anticipate that they will find it too long, and that some of the items can be removed to make it more feasible to apply. We also expect that the panel will be able to provide comments on whether some items need rewording or clarification, and possibly suggest some new/alternative items. Although we anticipate relatively minor feedback at this point, we want to leave it open to the panel to suggest major changes as well, and have thus incorporated this possibility in our methods.

The evaluation and refinement will be conducted during three rounds following a procedure inspired by the Delphi panel method (Whiting et al., 2017).

Round 1: Overall Written Assessment, Rating of Importance and Clarity, and Proposal of Alterations and Additions

In Round 1, all expert panelists will read through the checklist in full and individually write an overall statement that takes into account a) the exhaustiveness of the checklist, b) the proposed domains and subdomains, and c) the feasibility in applying the checklist as part of a systematic review. This statement should be a short paragraph for each category, for example: “This checklist is not exhaustive as it misses out on important issues that are central to research integrity...”, or “This checklist could feasibly be conducted as part of a systematic review, as I believe it will take less than 1 hour to complete”. The panelists are asked to focus on the checklist as a whole, rather than individual items, and to holistically consider the domains and subdomains.

Having written their general statement, they will then consider each item separately, and rate them on a scale from 1 to 5 on the importance of staying on the checklist (1 = “Not important”, 5 = “Fully important”). The panelists will also be asked to rate on a scale from 1 to 5 how clearly each item is worded (1 = “Not clear”, 5 = “Fully clear”).

In the final part of this round, the panelists are given the opportunity to suggest rewordings of specific items and propose additional items to the checklist. New items can be proposed both as a fully worded item, or as a general idea (e.g., “Include something about x”). Rewordings can be suggested both as a fully rewritten item or as a general statement (e.g., “It is not clear what x means in this item”). All verdicts, ratings, new items, and item rewordings by the panel are noted in an assigned spreadsheet. This spreadsheet also contains the exact wordings of the questions asked to the panel. It is available in the Supplementary Material [NOTE: will be added at Stage 2 of the Registered Report].

At the end of Round 1, after the written statements, ratings of each item’s importance and clarity, and suggestions for alterations and additions have been collected, the Core Team will refine the checklist accordingly. First, the Core Team will consider the overall verdicts qualitatively and jointly, as if they were peer-review statements for a submitted article. This

will inform potential changes to areas and subdomains, and any other overarching comments about the checklist as a whole.

Next, the Core Team will calculate and record the mean rating for each item for both importance and clarity, along with its standard deviation. We will also visually inspect the distribution for each item. All items with an average importance rating of <2 will be immediately discarded from the checklist. For the remaining items, we will inspect the ratings for clarity and reword all items that are rated as lower than 4, incorporating any suggested rewordings from the panel. After the refinement of the checklist, it will be distributed back to the Expert Panel.

Round 2: Evaluation of Changes, Additions, and Low Importance Ratings

In Round 2, the Expert Panel will individually evaluate all reworded items, the new proposed items, as well as any items with an average importance rating of 2 or 3 (even if not reworded). Panelists will be informed that they are now not rating the full checklist, but only items where additional ratings are needed. This time the panelists will simply vote to include or exclude an item. For each item where a simple majority (3 out of 5 panel members) voted for exclusion, the Core Team will remove the item.

The remaining items (importance ≥ 4 in Round 1 or a majority vote of “include” in Round 2) will comprise the proposed checklist for Round 3. The Core Team will at this stage develop the proposed scoring method. At the time of writing, our plan is to use the same scoring system as INSPECT-SR (a judgment per domain and an overall judgment; Wilkinson et al., 2024), but depending on the development of the item list during Rounds 1 and 2 (as well as feedback from the panel), this might have to change.

Round 3: Panel Roundtable Discussion

In Round 3, the Expert Panel will carefully read the proposed checklist, and then discuss it in a Zoom meeting with all panelists present. They will also discuss the proposed scoring method. This meeting will be moderated by at least one member of the Core Team. During this meeting, the alpha version of the checklist will be decided on, and a joint statement following the same format as in Round 1 on it will be made. Provided that this joint statement is positive, the checklist will move forward to Phase 3 (piloting of the checklist by the User Panel, see Figure 1). If a positive consensus statement cannot be reached at this point, the expert panel will instead jointly propose a way forward for revising the checklist. If this is the case, Core Team revisions and Round 3 may be repeated until positive consensus is reached by the panel (see dotted arrow in Figure 1).

Analysis of the checklist properties

We will provide descriptive statistics of the survey responses, including the mean and SD of each scored item and plotted distributions.

Phase 3: Piloting of the Checklist by the User Panel

Inclusion Criteria for Test Papers

The set of test papers that will be distributed to the User Panel originates from a scoping review of school-based educational interventions for students with intellectual disability (Batinović et al., 2025). A detailed description of eligibility criteria is available in the scoping review.

To compile the sample of studies sent to the User Panel raters, we will stratify the set of studies based on study design, and randomly select a study from each of the following strata: a randomized control trial, quasi-experimental group design, and a single-case experimental design. This way, each rater will evaluate one study per study design. We will assign each rater three papers in total to evaluate. All raters will evaluate the same three papers.

Piloting Procedure

Each member of the end-user panel will evaluate the papers using the checklist, and then rate the perceived usability of the checklist. The following steps will be requested from each member:

- Answer each checklist item for the assigned papers, one paper at a time;
- Time the evaluation process;
- Assess the usability of each item after evaluating all three papers;
- Assess the feasibility of each item after evaluating all three papers;
- Note down any additional feedback in an open-ended comment.

Users will evaluate the studies according to the alpha version of the checklist (finalized in Phase 2) in a spreadsheet. A proposed structure of the spreadsheet is available on [OSF](#).

Once a panelist has finished the assessment of all three papers, they are asked to rate each checklist item on its usability and its feasibility. In their spreadsheet, usability will be defined as the extent to which the rater finds the checklist items clear enough to implement, important to assess in the context of systematic reviews, and relevant to the content of the evaluated studies. The panelists will rate each item's usability on a scale from 1 to 5 (1 = "Not usable", 5 = "Fully usable").

Feasibility will be defined as the amount of time and additional effort required to answer an item confidently (e.g., use of external tools or search engines, information not provided in the text, difficult to implement on a certain study type). Time will be noted in minutes for each evaluated study, and will cover the time from the beginning of reading the paper, until they finish answering the last item on the checklist. Feasibility will be scored from 1 to 5 (1 = "Not feasible", 5 = "Fully feasible") for each item.

As suggested by Whiting et al. (2017), if deemed necessary, we will conduct interactive piloting sessions online (as a more feasible alternative to in-person hackathons) to facilitate feedback on the procedure and the tool.

After the User Panel evaluation is completed, the Core Team will refine the checklist based on comments and suggestions, if necessary. In case there are substantial changes to the number of items or the content of the checklist, we will repeat the testing again (drawing a new random stratified sample from the test papers).

Depending on the amount of changes and levels of agreement (both within the User Panel and within the Core Group), the Core Group may confer with the Expert Panel from Phase 2 before finalizing and publishing the QUERIES-Edu checklist.

Data Analysis

We will present descriptive statistics for each item for each of the evaluated designs, i.e., mean, SD, and range of the score and feasibility of each item for each evaluated study. This analysis pertains to data from the User Panel, while the feedback notes from the Expert Panel will be available on the OSF as supplementary material.

Ethical Considerations

There are no ethical considerations necessary for this project, as it does not involve human test subjects or any collection of personal information. While we survey panelists, they are participating and active researchers who will be co-authors of the final article. The text appraisals requested from the User Panel will be conducted on published (publically available) research papers only.

Expected Outcomes

We plan to publish a piloted, user-friendly checklist that expedites the systematic review process and increases trustworthiness in the included studies. The QUERIES-Edu checklist and accompanying guide will have the following features:

- Description of each item and detailed guidelines for answering them;
- A scoring system based on the results;
- Recommendations for integrating the checklist into systematic review workflows

Dissemination and Future Directions

The QUERIES-Edu checklist and the accompanying guide will be published open-access and be free to use.

In the future, we plan to build software that facilitates simultaneous assessment of multiple studies, as an addition to the checklist. The checklist will be available on a website created to simplify updating of the checklist and additional features.

Author Contributions [NOTE: will be updated at Stage 2 of the Registered Report to add contributions in investigation, data curation etc.].

CrediT statement created with the Tenzing app.

Conceptualization: L.B.¹, M.T.¹, N.H.¹, H.D.¹, and R.C.¹

Methodology: L.B., M.T., N.H., H.D., and R.C.

Project administration: L.B.

Supervision: M.T., H.D., and R.C.

Visualization: N.H.

Writing - original draft: L.B., M.T., N.H., H.D., and R.C.

Writing - review & editing: A.F.², J.H.², G.N.², S.J.W.², J.C.², N.J.B.², E.T.³, L.C.³, T.N.³, and A.K.³

¹Core Team

²Expert Panel

³User Panel

References

- Abalkina, A., & Bishop, D. (2023). Paper mills: A novel form of publishing malpractice affecting psychology. *Meta-Psychology*, 7. <https://doi.org/10.15626/MP.2022.3422>
- Batinović, L., Topor, M., Henriksson, I., Allgulin, M., Jonson, E., Carlsson, R., & Danielsson, H. (2025). *School-based interventions for primary and secondary school students with intellectual disability: A scoping review*. PsyArXiv. https://doi.org/10.31234/osf.io/heujm_v1
- Beall, J. (2012). Predatory publishers are corrupting open access. *Nature*, 489(7415), 179–179. <https://doi.org/10.1038/489179a>
- Boughton, S. L., Wilkinson, J., & Bero, L. (2021). When beauty is but skin deep: Dealing with problematic studies in systematic reviews. *Cochrane Database of Systematic Reviews*, 6. <https://doi.org/10.1002/14651858.ED000152>
- Chambers, C. D., & Tzavella, L. (2022). The past, present and future of Registered Reports. *Nature Human Behaviour*, 6(1), 29–42. <https://doi.org/10.1038/s41562-021-01193-7>
- Cochrane. (2021). *Editorial policies | Cochrane Library*. <https://www-cochranelibrary-com.e.bibl.liu.se/cdsr/editorial-policies>
- Else, H. (2024). ‘Systematic reviews’ that aim to extract broad conclusions from many studies are in peril. <https://doi.org/10.1126/science.zpniyp6>
- Else, H., & Van Noorden, R. (2021). The fight against fake-paper factories that churn out sham science. *Nature*, 591(7851), 516–519. <https://doi.org/10.1038/d41586-021-00733-5>
- Frandsen, T. F. (2019). How can a questionable journal be identified: Frameworks and checklists. *Learned Publishing*, 32(3), 221–226. <https://doi.org/10.1002/leap.1230>
- Gersten, R., Fuchs, L. S., Compton, D. L., Coyne, M. D., Greenwood, C. R., & Innocenti, M. S. (2005). Quality Indicators for Group Experimental and Quasi-Experimental Research in Special Education: *Exceptional Children*, 71(2), 149–164. <https://doi.org/10.1177/001440290507100202>
- Grey, A., Bolland, M. J., Avenell, A., Klein, A. A., & Gunsalus, C. K. (2020). Check for publication integrity before misconduct. *Nature*, 577(7789), 167–169. <https://doi.org/10.1038/d41586-019-03959-6>
- Heathers, J. (2024). *Approximate 1 in 7 Scientific Papers Are Fake v1.pdf*. Open Science Framework. <https://osf.io/https://osf.io/23zcr>
- Héroux, M. E., Butler, A. A., Cashin, A. G., McCaughey, E. J., Affleck, A. J., Green, M. A., Cartwright, A., Jones, M., Kiely, K. M., van Schooten, K. S., Menant, J. C., Weweg, M., & Gandevia, S. C. (2022). Quality Output Checklist and Content Assessment (QuOCCA): A new tool for assessing research quality and reproducibility. *BMJ Open*, 12(9), e060976. <https://doi.org/10.1136/bmjopen-2022-060976>
- Hyltse, N., Carlsson, R., & Nilsson, G. (2023). *Unconference Workshop: Risk Of Significant Error (ROSE) A checklist to assess the risk of errors in statistical reporting and datasets*. SIPS 2023, Padova.

- Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, 2(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science*, 23(5), 524–532. <https://doi.org/10.1177/0956797611430953>
- Lakens, D., Adolphi, F. G., Albers, C. J., Anvari, F., Apps, M. A. J., Argamon, S. E., Baguley, T., Becker, R. B., Benning, S. D., Bradford, D. E., Buchanan, E. M., Caldwell, A. R., Van Calster, B., Carlsson, R., Chen, S.-C., Chung, B., Colling, L. J., Collins, G. S., Crook, Z., ... Zwaan, R. A. (2018). Justify your alpha. *Nature Human Behaviour*, 2(3), 168–171. <https://doi.org/10.1038/s41562-018-0311-x>
- Lakens, D., Hilgard, J., & Staaks, J. (2016). On the reproducibility of meta-analyses: Six practical recommendations. *BMC Psychology*, 4(1), Article 1. <https://doi.org/10.1186/s40359-016-0126-3>
- Liverpool, L. (2023). AI intensifies fight against ‘paper mills’ that churn out fake research. *Nature*, 618(7964), 222–223. <https://doi.org/10.1038/d41586-023-01780-w>
- Mol, B. W., Lai, S., Rahim, A., Bordewijk, E. M., Wang, R., van Eekelen, R., Gurrin, L. C., Thornton, J. G., van Wely, M., & Li, W. (2023). Checklist to assess Trustworthiness in RAndomised Controlled Trials (TRACT checklist): Concept proposal and pilot. *Research Integrity and Peer Review*, 8(1), 6. <https://doi.org/10.1186/s41073-023-00130-8>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Rele, S., Kennedy, M., & Blas, N. (2017). Journal Evaluation Tool. *Librarian Publications & Presentations*. https://digitalcommons.lmu.edu/librarian_pubs/40
- 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>
- Sotola, L. K. (2022). Garbage In, Garbage Out? Evaluating the Evidentiary Value of Published Meta-analyses Using Z-Curve Analysis. *Collabra: Psychology*, 8(1), 32571. <https://doi.org/10.1525/collabra.32571>
- Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., ... Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ*, 14898. <https://doi.org/10.1136/bmj.14898>
- Whiting, P., Wolff, R., Mallett, S., Simera, I., & Savović, J. (2017). A proposed framework for developing quality assessment tools. *Systematic Reviews*, 6(1), 204. <https://doi.org/10.1186/s13643-017-0604-6>

Wilkinson, J., Heal, C., Antoniou, G. A., Flemyng, E., Alfirevic, Z., Avenell, A., Barbour, G., Brown, N. J. L., Carlisle, J., Clarke, M., Dicker, P., Dumville, J. C., Grey, A., Grohmann, S., Gurrin, L., Hayden, J. A., Heathers, J., Hunter, K. E., Lasserson, T., ... Kirkham, J. J. (2024). Protocol for the development of a tool (INSPECT-SR) to identify problematic randomised controlled trials in systematic reviews of health interventions. *BMJ Open*, 14(3), e084164. <https://doi.org/10.1136/bmjopen-2024-084164>

Appendix 1.

List of Candidate Checklist Items

Domain: Journal/Publisher

1. Is the journal indexed in databases (e.g., DOAJ, Web of Science, PubMed)?

Guide: Search for the journal name in [Ulrichsweb](#) to locate all directories in which the journal is indexed, although this site requires institutional access. You can also use the [Norwegian Register](#) of Journals to see if the journal is indexed in the Directory of Open Access Journals (DOAJ). Clarivate's Web of Science and PubMed are better indicators, as DOAJ does not capture journals which are not open access. An open source alternative could be [OpenAlex](#), however it is unclear what the database currently covers, which means that it might not capture certain journals.

2. Does the journal have a clear conflicts of interest policy for editors, authors, and reviewers?

Guide: This information should be easily findable, most often in the submission guidelines or journal overview sections. Look for dedicated sections outlining the journal's policies on conflicts of interest.

3. Is there clear information about copyright or licensing?

Guide: Information about copyright or licensing should be easily findable, most often in the submission guidelines or journal overview sections. Look for details regarding how authors can use their work and what rights they retain.

4. Is the time between submission and acceptance reasonable?

Guide: Look for typical timelines for the review process provided by the journal. Note that extremely short editorial processes (i.e., a couple of days between submission and acceptance or review to revision) are a red flag. Although there could be legitimate reasons for a fast editorial process (e.g., streamlined peer review), in connection to other concerns this might be an indicator of a predatory journal.

5. Is there a clear peer-review policy available on the journal website?

Guide: Information about peer review and the journal's review policy should be easily findable, most often in the submission guidelines or journal overview sections. There should be information on what type of peer review is conducted (e.g., double blind) and whether the reviews are open and available for published articles. Types of peer review do not indicate more or less trustworthy journals, but a red flag is missing information about peer review practices. *This should be assessed as a general practice, and does not ensure that each article has been through a proper, unbiased peer-review, which may or may not be better indicated in the article itself (e.g., if the reviewers are named or peer-review reports shared).*

6. Is the journal's scope/aims section too broad or vague?

Guide: The scope or name of the journal should provide this information. Journals with very broad scopes/aims or titles including terms like "world," "global," or "international" might warrant further checks of the scope to ensure specificity and relevance, particularly tied to the article that is evaluated.

Appendix 1.

List of Candidate Checklist Items

7. Does the journal charge multiple fees for submission, publication, handling or withdrawal or offers fast track handling for a fee?

<i>Guide:</i> This information should be easily findable, most often in the submission guidelines, journal overview sections, or on the publisher's website. Multiple fees, especially ones an author has to pay upon submission, indicate predatory practices. If a journal provides fast-track options for a specific fee or charges withdrawal, that can also indicate problematic practices.

8. Does the journal have DOIs?
--

<i>Guide:</i> Check if the articles published by the journal have Digital Object Identifiers (DOIs). This information should be visible in the articles themselves or mentioned on the journal's website. Although this is an indicator of good practices, older articles will not have DOIs, and newly-founded or community-run journals might not be able to provide this service due to financial or infrastructural reasons. However, unless it is an old article, even the potential financial reasons do not warrant clearance and this item can be considered concerning if answered “no”.
--

9. Does the journal use unconventional payment methods?

<i>Guide:</i> Information about payment methods should be easily findable, most often in the submission guidelines or journal overview sections, or redirected to the information provided by the publisher. Unconventional or suspicious payment methods are highly concerning practices, e.g., requiring payment through PayPal.

10. Is manuscript submission done via email or third-party contact form?
--

<i>Guide:</i> The method for manuscript submission should be clearly stated, most often in the submission guidelines or journal overview sections. Avoid journals that only accept submissions via email or third-party contact forms without a secure submission system.
--

11. Is the website poorly laid out?

<i>Guide:</i> Evaluate the overall design and functionality of the journal's website, check whether there are inconsistent fonts or a non-cohesive structure of the website.

12. Is the website secure (no security warnings or unsafe connection indications)?
--

<i>Guide:</i> Check for security features on the website, such as a secure connection (https://) and the absence of security warnings. A secure website indicates better handling of users' data and submissions.

13. Does the journal mention "fast" or "easy" submission?

<i>Guide:</i> Look for terms like "fast" or "easy" submission in the submission guidelines or journal overview sections.

14. Is there contact or affiliation information for board members?
--

<i>Guide:</i> Information about the editorial board, including contact or affiliation details, should be available in the section about the editorial board.

Appendix 1.
List of Candidate Checklist Items

15. Is there access to previous articles (even if only to the title/author information of the article)?

Guide: The journal's website should provide access to all published issues and titles of articles published in those issues. Ensure you can view previous publications, even if full-text access is restricted. Difficult access to archive issues indicates issues with the credibility of the journal.

16. Is publisher contact information available and verifiable?

Guide: The journal overview section or the journal's website should clearly identify the publisher. If not, check OpenAlex, the Norwegian registry or Ulrich's web to verify the publisher's contact information.

17. Does it appear the journal was previously owned by a different publisher, and is the integrity of the previous publisher verifiable?

Guide: Check the journal's ISSN record at <https://portal.issn.org> to find information about previous publishers and verify the existence of the journal. It is not uncommon for journals to switch publishers, but understanding whether journals belonged to untrustworthy publishers at some point could indicate potential issues for articles published at that time.

18. Is the editorial board related to the topic of the journal?

Guide: Check the editorial board members' backgrounds and affiliations to ensure they are relevant to the journal's subject matter. International or interdisciplinary expertise does not have to be concerning if the scope of the journal is broad, but unclear connection between the members' background and journal's scope is a red flag.

19. Is there evidence of email spamming?

Guide: Search for the journal name along with terms like "spam" in your search engine to see if there are reports of the journal engaging in email spamming. Especially if the journals invite authors to publish their work in the journals, or invite reviewers who do not match the topic of the article.

20. Is there access to the full text of published articles?

Guide: The journal should provide access to all published articles, with a full-text version available through subscription, payment, or library access. Verify that it is possible to access the full text, e.g., by contacting the librarian at your institution.

Appendix 1.

List of Candidate Checklist Items

Domain: Article

1. Is there a statement about ethical approval and consent?

Guide: This information should usually be available at the end of the article or somewhere in the methods, under ethical consideration headings. Otherwise, look for information in the description of the participants or procedure. Properly, if applicable, this information should state who the institutional board was (e.g., name of the university), and the authors should provide a reference number for the approved application. Furthermore, there should be stated compliance with the Helsinki declaration. Stated compliance with the Helsinki declaration without clear statement of who approved the application should be taken with caution, as this is too vague and does not indicate received ethical approval per se. There could also be commercial review institutions providing ethical approval, which should again be taken cautiously, as commercial institutions have financial incentive to provide approval. A statement could also indicate that ethical approval was not required, but no statement is problematic.

2. Are there tortured phrases (synonyms that do not comply with standard terminology)?

Guide: Look for potential terms that are nonsensical or emulate conventional terminology but use synonyms that are not common or inappropriate in the context of the paper. More guidance can be found here: <https://arxiv.org/abs/2107.06751> and <https://arxiv.org/abs/2210.04895>. Use of tortured phrases can indicate plagiarism or inappropriate use of AI.

“The most promising method is based on large language model token predictions propagate to their noun chunks. It achieves a good recall (0.87) but the precision still needs to be improved (0.61). This means that the detection of tortured phrases still requires some sort of manual checking by domain experts. We also noticed that distinguishing tortured phrases from their legit counterpart can be highly contextual.”

From: <https://arxiv.org/html/2402.03370v1#bib.bib8>

“Examples of improbable word sequences extracted from the SCIGen grammar include:

- “in fact, few futurists would disagree with”.
- “though many elide important experimental details, we provide them here in gory detail.”
- “A well designed system that has bad performance is of no use to any man, woman or animal.”
- Featuring typos:
 - “but without all the unnecessary complexity.”
 - “holds suprising results for patient reader.””

From: <https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/asi.24495>

“E.g. counterfeit consciousness instead of artificial intelligence.”

From: <https://aclanthology.org/2022.sdp-1.4.pdf>

Appendix 1.
List of Candidate Checklist Items

3. Is there a short list of references or are references missing?

Guide: Is the reference list too short for the size of the study, and are all in-text citations in the reference list and vice versa? Based on your expertise, are crucial seminal studies missing without justification?

4. Are there large sections of text that are nonsensical or incoherent?

Guide: Are there miscitations (e.g., citing studies that are not relevant, like citing a study on autism in a different context), or chunks of text that contain incoherent wording, missing words, or content that has nothing to do with the aim/topic of the study. While some grammatical errors and typos can easily slip past peer review and editorial process, this item refers to the content of the text rather than the (mis)use of language. Much nonsensical text can be attributed to “tortured phrases” (see item above) and other signs of automated/AI-generated text, but also direct contradictions within the writing (e.g., making one argument in the introduction and a complete counter-argument in the conclusion).

5. Is there enough information reported to assess the plausibility of the study (e.g., country, participant demographic information, study conductor, procedure description)?

Guide: Reporting standards for specific study designs could be helpful here, and one can search for appropriate guidelines based on study design through EQUATOR: <https://www.equator-network.org/reporting-guidelines/>. Missing information about the procedure is a red flag as it prevents a proper evaluation of the results and can hide potential biases.

6. Could the study plausibly be completed as described?

Guide: Based on information available in the paper (guided by item nr. 5), is it plausible that the intervention was implemented? For example, look for the type of population in relation to the sample size (e.g., is it possible to have 400 participants with intellectual disability in a small country), duration of data collection (finished in 3 days for 400 participants), number of authors/contributors involved, affiliation of researchers in relation to the location of the study.

7. Are tables and figures of sufficient quality to inspect them properly (e.g., not just screenshots of statistical software)?

Guide: Look at the quality of the images and legibility of the tables and values on graphs, as interpretation of results can be impeded by improper formatting of tables, or low resolution and distorted images and graphs.

8. Are any reporting standards specified?

Guide: This information should be at the start of the Methods section, and the authors should state/cite which reporting guidelines they followed.

Appendix 1.
List of Candidate Checklist Items

9. Are there any discrepancies between data reported in figures, tables and text?

Guide: Look for any inconsistencies in numbers, colors, labels, or measurement types. Is the number of participants/observations consistent within the manuscript (or if there is attrition is that clearly stated)? Are reported numbers within reasonable/possible ranges (e.g., a mean of 7.4 for an item that can be scored between 1-5)? When applicable, do numbers add up as expected (for instance the mean values of two subgroups and the overall mean value, or do percentages add up to a 100 where applicable)? Are labels consistent for all groups (e.g., swapped colors on two independent figures)?

10. Is there evidence of copied work within paper, such as duplicated or partially duplicated tables or figures/images?

Guide: Do multiple tables seem to have identical values or partially identical values without proper justification for it? Do figures (or images) seem duplicated without this being justified anywhere in the text/supplemental materials?

11. Does the author list make sense given the nature of the study?

Guide: For example, does a simple study have dozens of authors from different institutions and with diverse expertise? Look for appropriateness of their affiliations/expertise, and necessity of the number of authors given the type of a study. For example, would a single-case design study require 30 authors, and what were their roles?

12. Is there a clear conflict of interest statement?

Guide: Look for a COI statement and whether possible conflicts have been properly addressed, for example, one of the authors is affiliated with a company that could benefit from the study? This requires looking into authors and evaluating potential collaborations that could cause a COI.

13. Was the time between submission to acceptance reasonable?

Guide: If stated in the paper, how long did it take from submission to acceptance? A couple of days (less than a week) is potentially problematic, unless there is clear justification (e.g., streamlined peer-review, resubmission categorized as a new submission etc). Generally, it is concerning when the paper has a drastically shorter acceptance time than the average time the journal states this process lasts.

14. Has the study been prospectively registered?

Guide: If the study links or states a preregistration link or a trial registration number, check if these were registered after the study was conducted. It is particularly problematic if the authors do not explicitly state that the registration happened after the study was conducted.

15. Are details such as dates and study methods in the publication consistent with those in the registration documents?

Guide: Check consistency of the content between the registration and the published paper.

Appendix 1.
List of Candidate Checklist Items

16. Has the study been retracted or does it have an expression of concern, a relevant post-publication amendment, a critical Retraction Watch or PubPeer comment or has been previously excluded from a systematic review?

Guide: Check to see whether this study has an expression of concern or retraction notice on the journal website, or relevant critical comments on Retraction Watch

(<http://retractiondatabase.org/RetractionSearch.aspx?>) or PubPeer (<https://pubpeer.com/>).

Zotero reference manager provides add-ons which can facilitate simultaneous checking multiple papers for retraction/PubPeer comments. <https://www.zotero.org/support/plugins>

Appendix 2.

Instructions for the Expert Panel

Expert panel

Overall, the items presented are deemed important, but not necessarily feasible, and the core team acknowledges some items could fall out of scope regardless of relevance and importance. The expert panel should therefore consider the wording of the items, feasibility of answering each item, and weigh it against the potential benefit of including potentially longer items.

One should note that items for the journal/publisher assessment should be evaluated slightly differently than the article items, as it is likely that one journal assessment could apply to multiple articles, thus shortening the overall evaluation time of the entire set of papers.

You are asked to evaluate the first set of items for the research integrity checklist for evaluation of special education studies. Begin by looking through the entire attached checklist to familiarize yourself with the type of items and its structure.

Your first task is to provide a written statement describing your general opinion of the items and the checklist as a whole. You should include an opinion about the following in the written statement:

- a) How exhaustive is the checklist?
- b) Are the proposed areas and subdomains relevant/appropriate?
- c) How feasible is it to apply this checklist during a systematic literature review to exclude untrustworthy studies?

You can think of this written statement as if you are a peer reviewer for this proposed scale, and you are free to include other points besides the three mandatory ones.

After you complete the general statement which answers the questions and includes your general opinion, score each individual item based on clarity and importance in the provided spreadsheet. Both relevance and clarity are scored on a 1-5 scale, with 1 being not clear/important and 5 being fully clear/important. The first sheet includes items for the Journal domain, and the second sheet includes items for the Article domain.