Proposal Narrative

Developing Web-tools to Verify the Transparency and Replicability of Education Research

Description of Project

A central problem in education research is determining which interventions work, in the sense that they actually produce the desired learning outcomes. Like in most other empirical sciences, the current state-of-the-art strategy is to systematically review (and sometimes meta-analyze) evidence from all known studies of an educational intervention (as is used by the US Department of Education's What Works Clearinghouse). Such strategy is a useful starting point, but is insufficient as a tool to soundly evaluate the effectiveness of educational interventions because the transparency and replicability of the included studies are not ensured.[1] Without sufficient transparency, the evidence for a finding cannot be taken at face value because undisclosed analytic and design flexibility may have been (un)intentionally exploited by researchers (John et al., 2012; Simmons et al., 2011). And without replicability (i.e., if an educational finding is not consistently observable in independent samples using similar methodologies), one cannot rule out that a finding was caused by various technical/experimenter errors, fraud, or chance. Hence without transparency and replicability, a study's evidence cannot be trusted.

It is difficult to overstate the importance of this problem. Because of it, education researchers (nor What Works Clearinghouse reviewers) cannot determine the actual

[1] Indeed, the words "transparency", "reporting standards/guidelines", and "replication" are not used once in the WWC's 130-page Standards Handbook.

efficacy of interventions in producing desired learning outcomes. This means, for example, that teachers are likely using techniques that will not actually improve learning (and that students are failing to benefit from alternative *effective techniques*). This also means that theoretical progress in the field of education research is being substantially retarded because it is not possible for education researchers to know which published findings can be productively built upon.

Indeed, there is a growing awareness in education research regarding the crucial need for increased transparency and replication studies to verify the accuracy and credibility of published education findings (Hedges & Schauer, 2019; Makel & Plucker, 2014a; Makel & Plucker, 2014b; Pashler, McDaniel, Rohrer, & Bjork, 2009). This awareness is part of a more general movement in the social sciences that is placing a higher emphasis on the transparency and replication of studies (e.g., OSC, 2015; Munafò et al., 2017; Nosek et al., 2015; list of 170+ open science initiatives).

This is an incredibly positive development. However, increased transparency and the regular execution of replication studies are still insufficient for assessing the credibility of a finding *without careful curation*. What is needed is an easy-to-access system (1) to ensure articles/studies comply with the appropriate transparency standards and (2) to track, organize, and interpret replications so that the replicability (and generalizability; see below) of effects/findings can be determined.

The goal of the project is to tackle education research's credibility problem by developing such an online curation system. In particular, we will develop new web tools for CurateScience.org, a next-generation multi-stakeholder transparency compliance

and accountability system for scientific articles (see Figure 1 below). The curation platform/tools will allow education researchers to identify educational interventions that can be trusted by providing tools to check (1) whether a study complies with an appropriate transparency standard and (2) whether it has been successfully independently replicated. The curation tools will substantially improve the validity of What Works Clearinghouse's systematic reviews of the efficacy of educational interventions, will accelerate development of educational theory, and ultimately will improve educational outcomes worldwide (see Figure 16 for additional benefits).

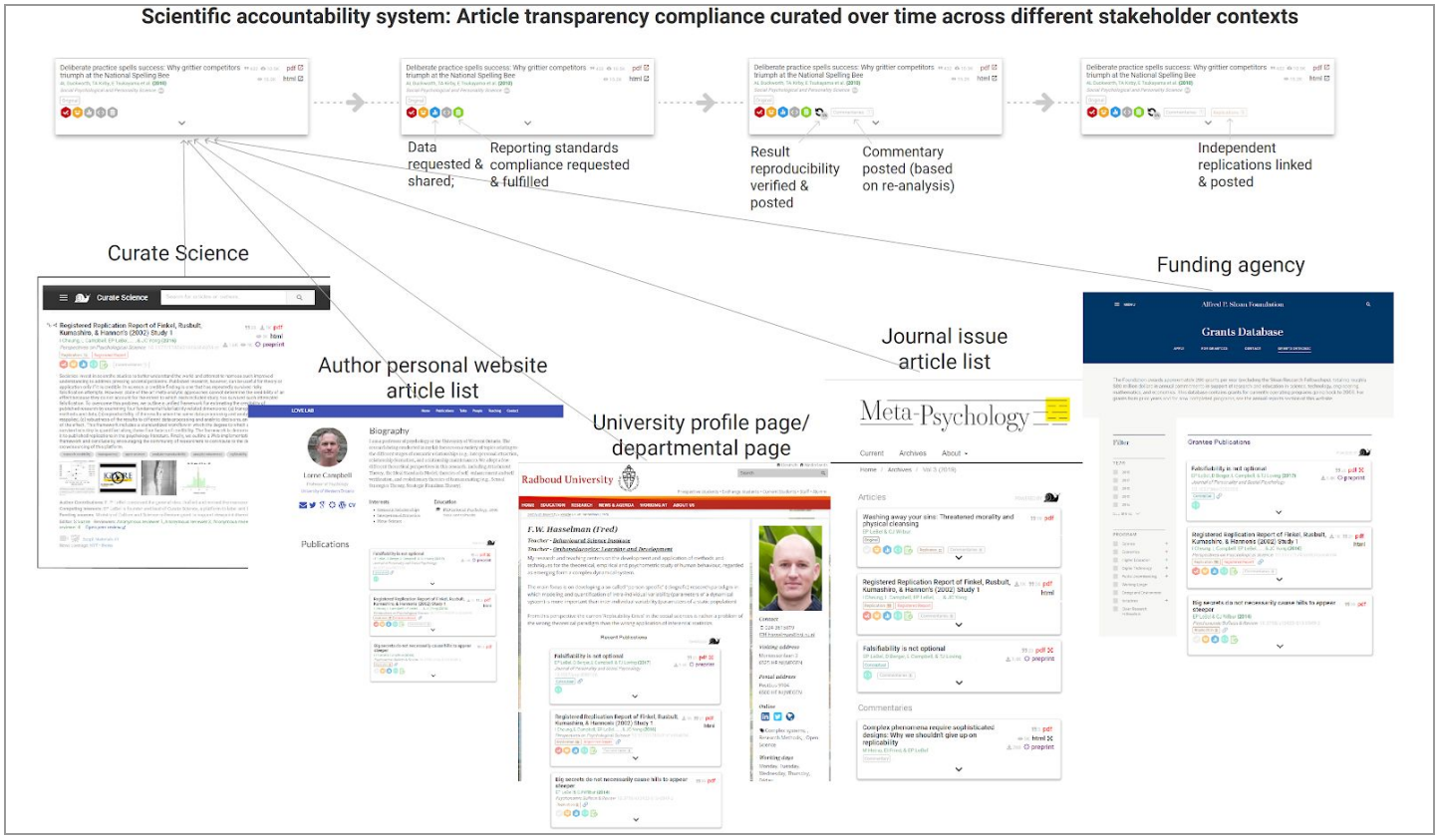


Figure 1. Proposed transparency compliance and accountability system: An article's transparency, result reproducibility, and effect replicability are verified by the research community over time. Findings that survive scrutiny are treated as credible proportional to the extent of exposure/scrutiny from peers achieved over that time period.

Project Rationale

General replicability problems in the social sciences are now well known (OSC, 2015; Munafò et al., 2017; Nosek et al., 2015; Pashler & Wagenmakers, 2012). Replicability issues specific to education research findings are also well documented. For example, replication difficulties have been reported for the following education-related findings:

- Stereotype threat on women's math performance (Finnigan & Corker, 2016; Inzlicht & Jensen, 2014; Stoet & Geary, 2012; Sunny et al., 2016),
- Mindset growth on scholastic achievement (Li & Bates, 2017; Bahnik & Vranka, 2017a),
- Grit on scholastic aptitude (Bahnik & Vranka, 2017b),
- Self-affirmation on scholastic aptitude (Hanselman et al., 2017),
- Music lessons/enrichment on cognitive aptitude (Mehr et al., 2013).

The proposed platform will allow education researchers to address its credibility problem head on. Non-contributing users will be able to track, organize, and interpret replication studies, yielding accurate estimates of the replicability and generalizability of educational interventions (more on generalizability below). Contributing users will also be able to curate the transparency compliance and replication of original/novel studies (with incentives to contribute built into the platform, see below). Transparency-compliant studies can then be used as a filter for What Works Clearinghouse systematic reviews. The platform will increase research efficiency by providing tools for education researchers to identify studies that have withstood

thorough scrutiny, and therefore are worth investigating more deeply. Curating transparency also acts as a preventive measure against low-quality research, which will increase research rigor/quality, and ultimately increase replicability.

To guide the design and implementation of the curation tools/platform, we have developed a harmonized framework to quantify the transparency, replicability, and generalizability of published findings in the social sciences (LeBel et al., 2018). The framework is based on *falsificationism* (Lakatos, 1970; Popper, 1959), the key principle underlying science's remarkable success as a knowledge production system. Under falsificationism, an idea must be tested in a way whereby the likelihood of proving the idea wrong is very high (if, in fact, the idea is incorrect). Testing (and reporting) an idea transparently increases the likelihood that other researchers can identify errors through follow-up re-analyses and replication. The more scrutiny an idea's evidence "survives", the more trustworthy/credible that evidence can be considered.

The framework is highly flexible. It can be applied to a wide range of article types (e.g., original, replication, meta-analysis, conceptual, simulation) and study methodologies (e.g., experimental, observational, longitudinal) from disparate social science fields including education, psychology, political science, economics, and sociology (each of which lacks a transparency and replication compliance system). The framework curates the following four properties:

- *Transparency*[2]:

[2] "Transparency" is used either in a narrow (primary/secondary transparency proper) or broad (umbrella term that includes transparency proper and reproducibility) sense, however, the intended meaning will be clear from the context.

    - *Primary transparency*: availability of study materials, underlying data/analysis code files, reporting guidelines compliance, and study/analysis preregistration.
    - *Secondary transparency:* author contributions, competing interests, funding sources, and peer-review information.
- *Result reproducibility*: reported results can be successfully reproduced by repeating the same statistical analyses on the original data.
- *Result robustness*: robustness of results to different data-analytic decisions.
- *Effect replicability*: effect can be consistently observed in new samples, at a magnitude similar to that originally reported, using methodologies/conditions similar to those of an original study.

For the purpose of this project, we will implement web-tools so that researchers can curate the (1) transparency, (2) result reproducibility, and (3) effect replicability of their own and other researchers' articles, at both the article- and study-levels (for feasibility reasons, result robustness is not part of the current project). For each curated article, transparency/replication information will be accessible, in an organized fashion, via expandable "article cards" (see Figure 2 below). These article cards, and any article card *lists*, will be available on CurateScience.org, but will also be displayable outside of the platform (e.g., on researcher's websites, university profile/department pages, journal issue article lists, etc.; see Figure 1). This affords scientific accountability given that any researcher will be able to add follow-up replications/commentaries to an article card across the different stakeholder contexts.

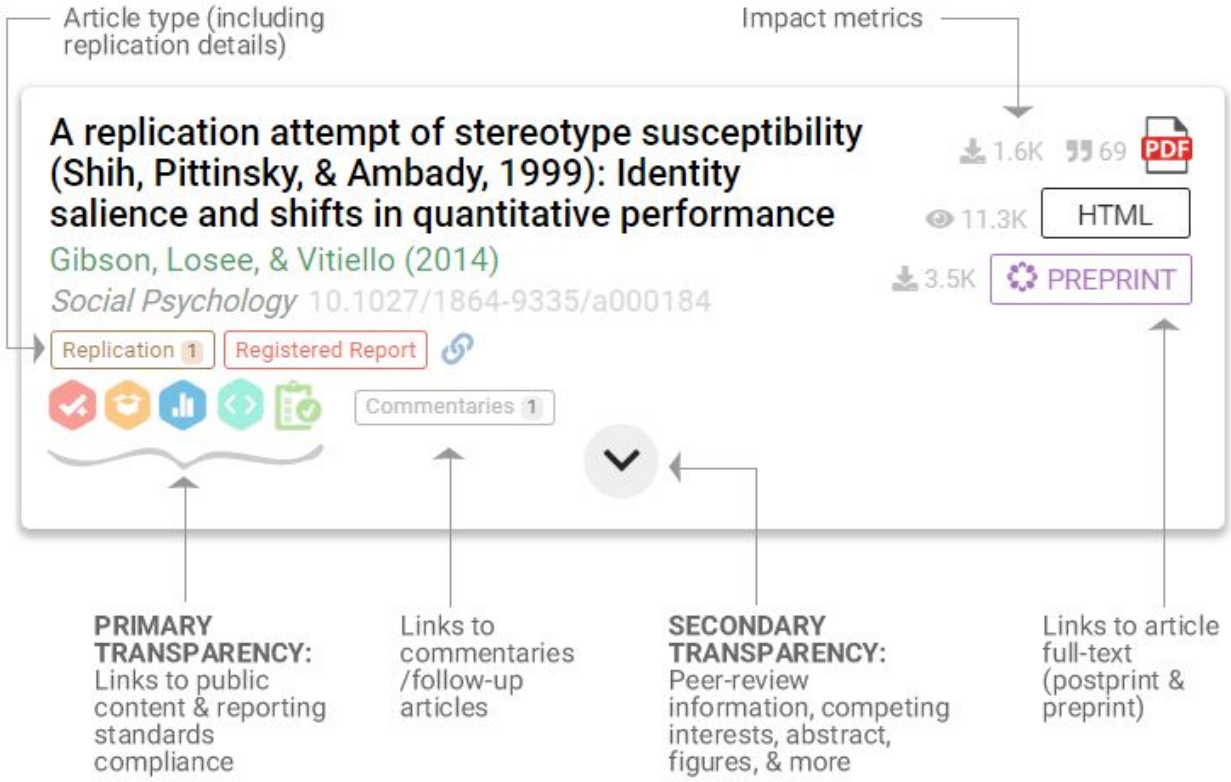


Figure 2: Basic layout of Curate Science's article card displaying information about an article's transparency, reproducibility, and follow-up replications/commentaries.

The framework and current prototypes are based on (and have been tested and refined by) curating more than N = 1,127 replication studies in psychology (see here), the largest known curation effort of its kind in the world. It provides an ideal starting point for curating studies in the educational sciences. In what follows, we describe the characteristics of the curation and compliance tools for transparency, reproducibility, and replicability/generalizability.

Transparency

We will develop transparency curation and compliance web-tools. The transparency curation web-tool will allow researchers to label and link a study's

transparency information. The transparency compliance web-tool can then be used to indicate whether the curated transparency information complies with a specific transparency standard.

Primary transparency curation. Primary transparency involves the five most fundamental kinds of transparency categories that correspond to the five most fundamental aspects of how a study is conducted and reported (see Figure 3 below; three of which were adapted from the Center for Open Science's open practice badges). That is:

1. *Preregistration*: Were (crucial aspects of) the planned study design, data processing choices, and statistical analyses preregistered at a recognized public registry *prior to data collection*?
2. *Open materials*: Are all experimental files, stimuli, and questionnaires needed to conduct a fair replication attempt publicly posted on an online repository? (or valid reason is given for non-availability)
3. *Open data*: Are the raw/processed data files needed to reproduce a study's primary substantive findings publicly posted on an online repository? (or valid reason is given for non-availability)
4. *Open code*: Are the syntax/code files used to execute the statistical analyses publicly posted on an online repository?
5. *Reporting standards compliance*: Are all necessary methodological details reported in an article according to official (methodology-specific) reporting guidelines?

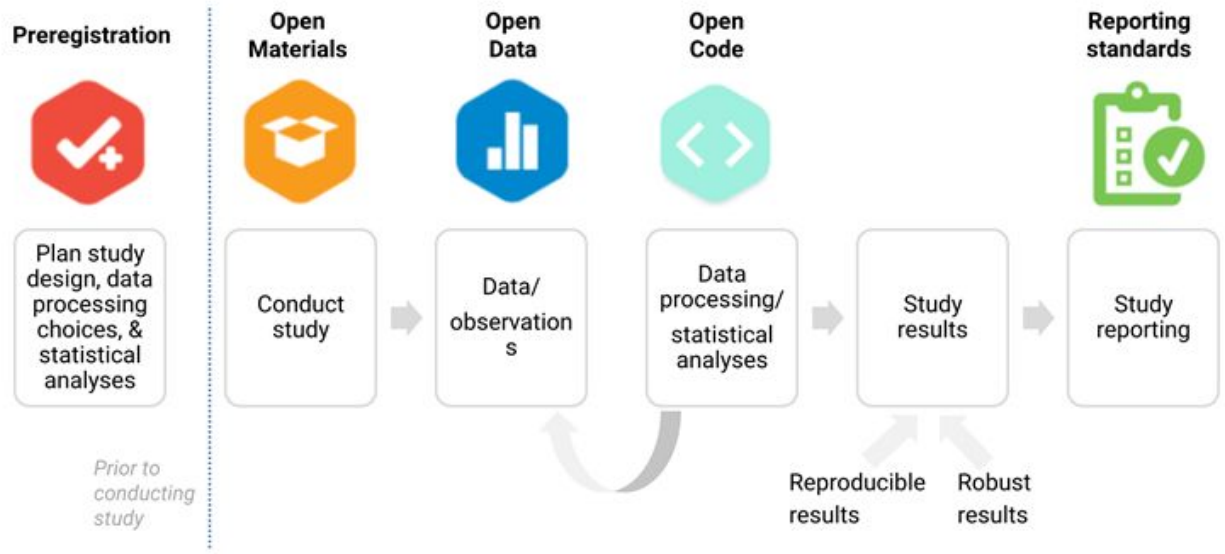


Figure 3: The five most fundamental kinds of transparency categories (primary transparency) that correspond to the five most fundamental aspects of how a study is conducted and reported.

Secondary transparency curation. Secondary transparency involves information about (if available):

1. *Author contributions*: Declaration of the substantive contribution of a paper's (co-)authors.
2. *Competing interests*: Declaration of interests that compete with one's interest to pursue the truth.
3. *Funding sources*: The name of all organisations that funded the research.
4. *Peer review information*: The name of the editor/action editor, peer reviewers, and/or URL to the actual reviews (a practice, growing in popularity, called *open peer review*).

These are crucial for researcher and journal accountability, and to properly place an article's finding in the broader (professional and funding) context in which it took place.

Transparency compliance. Having access to the primary and secondary transparency information allows one to ensure that an article *complies with* a specific transparency standard relevant to the article type and employed methodologies (e.g., see Figure 4, left panel).

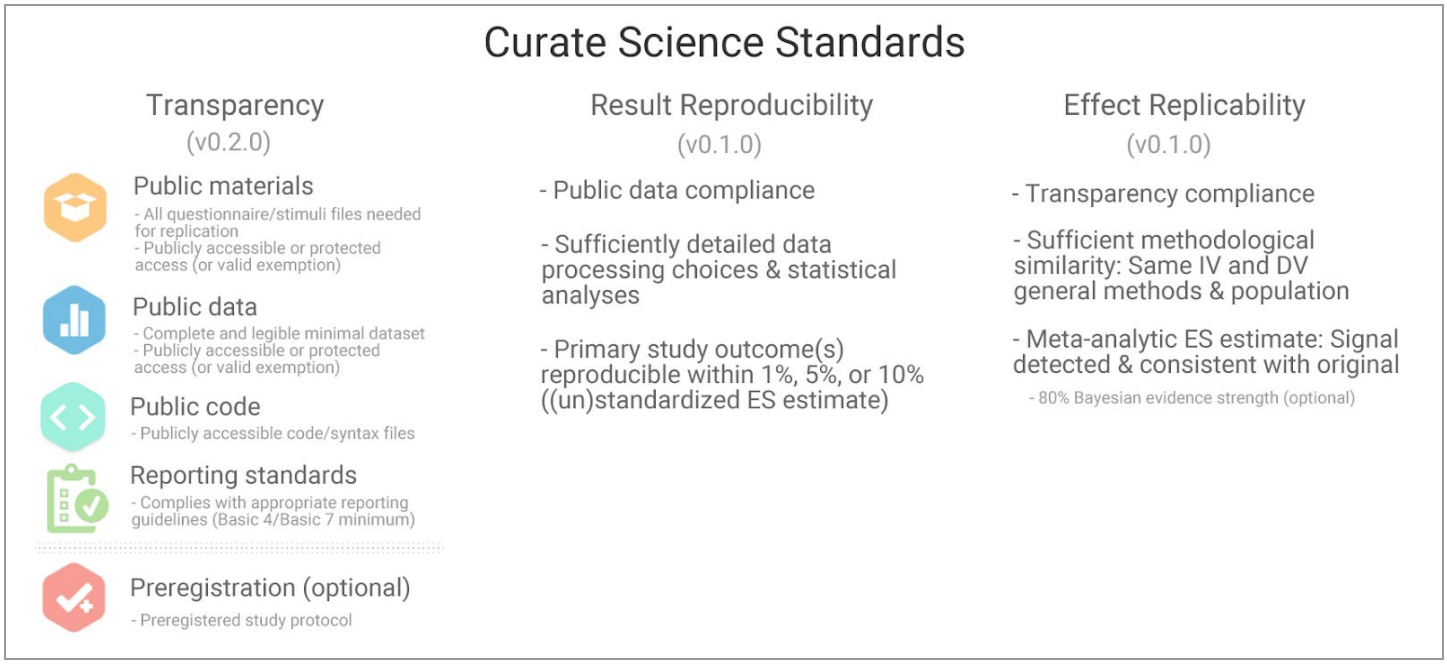


Figure 4. Requirements of proposed standards for transparency, result reproducibility, and effect replicability.

Our proposed transparency standard is intentionally more lenient than standards that are part of COS' TOP Guidelines (Nosek et al., 2015). This can be considered a "starter pack" to accelerate community uptake (see Figure 5 to compare requirements of these two standards) that will serve as a springboard to achieving higher standards (see Figure 5, right panel). Indeed, a menu of transparency standards will be available so researchers can choose a specific standard they desire (or need) to comply with, including journal- and funder-specific transparency standards (e.g., the journal *Meta-Psychology*'s transparency standard; see Figure 5).

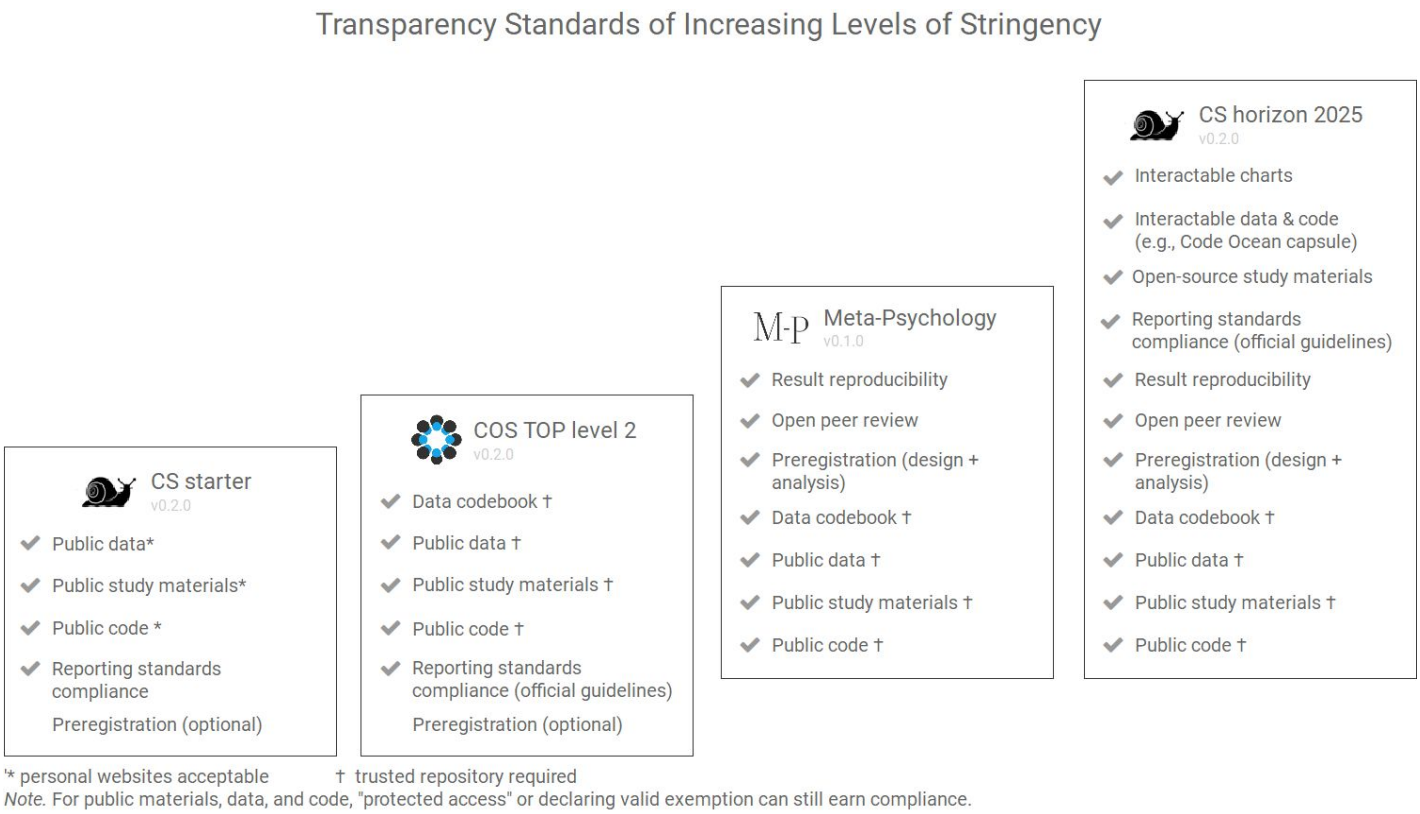


Figure 5. Transparency standards corresponding to higher levels of transparency given the increasing number of requirements needed to earn compliance (from left to right).

Result Reproducibility

Result reproducibility curation tool. The result reproducibility curation web-tool will allow an independent researcher to indicate whether reported results can be successfully reproduced by repeating an article's statistical analyses on the data. One will be able to do this for each study's primary outcome (or set of outcomes) that constitutes a study's "primary substantive" finding, which is defined as one which is emphasized in the article's title, abstract, figure, or table (Hardwicke et al., 2018).

Result reproducibility compliance tool. The independent researcher will then be able to indicate the reproducibility status of an article, according to a specific result reproducibility standard. For example, whether an article's reproducibility results comply

with our proposed result reproducibility standard (see Figure 4, middle panel), and which level of concordance was observed (e.g., primary study outcomes were reproduced within 5% of reported results).

Effect Replicability and Generalizability

We will also develop a replication curation/tracker tool and a replicability/generalizability compliance tool.

Replication curation/tracker tool. The first step in being able to verify the replicability of education intervention findings is to have a tool that identifies, curates, and tracks replication studies of education-related effects over time. This requires categorizing the methodological similarity of a follow-up study relative to a previously published study, which can be done using our replication taxonomy (LeBel et al., 2017). The taxonomy was specifically designed to allow categorization of the methodological similarity of a study (relative to a previous study) for a wide range of study types and methodologies (see Figure 11 below). Once replications are curated and linked to an original study, any new replications of that same original study (or of a related effect) can be tracked over time.

Replicability/generalizability compliance tool. Once (enough) replications are identified and curated, they can be organized into *replication collections*, so that an effect's replicability and generalizability can be estimated meta-analytically (according to a specific replicability standard, e.g., Figure 4, right panel). In such collections, replications are organized by distinct general methods of measuring an effect (and by distinct populations, e.g., high school vs. university population). In this way, one can

ensure that an effect is replicable *and* generalizable (i.e., an effect is consistently observable across distinct general methods and populations). For example, for the *longhand benefit effect* (Mueller & Oppenheimer, 2014), Figure 6 shows an idealized case of what it would look like for this effect to be replicable and generalizable. This scenario can be contrasted with two alternative scenarios: (1) an effect is replicable (for a specific method or population), but not generalizable to other methods/populations or (2) an effect is neither replicable nor generalizable.

In summary, curating the transparency, reproducibility, and replicability compliance of published studies/effects maximizes the probability that erroneous results introduced into the literature will be detected by the community. This will accelerate scientific progress in education research by allowing researchers to more rapidly (and efficiently) build upon solid/credible research findings that provide a strong foundation for the development of a sturdy knowledge structure.

Methods of Proposed Research

The proposed research involves mixed-methods drawn from the fields of computer science (product design, user experience/beta testing, and iterative/incremental development), archival/information science (manual curation, digital preservation, and information filtering), statistics, philosophy of science, and meta-research.

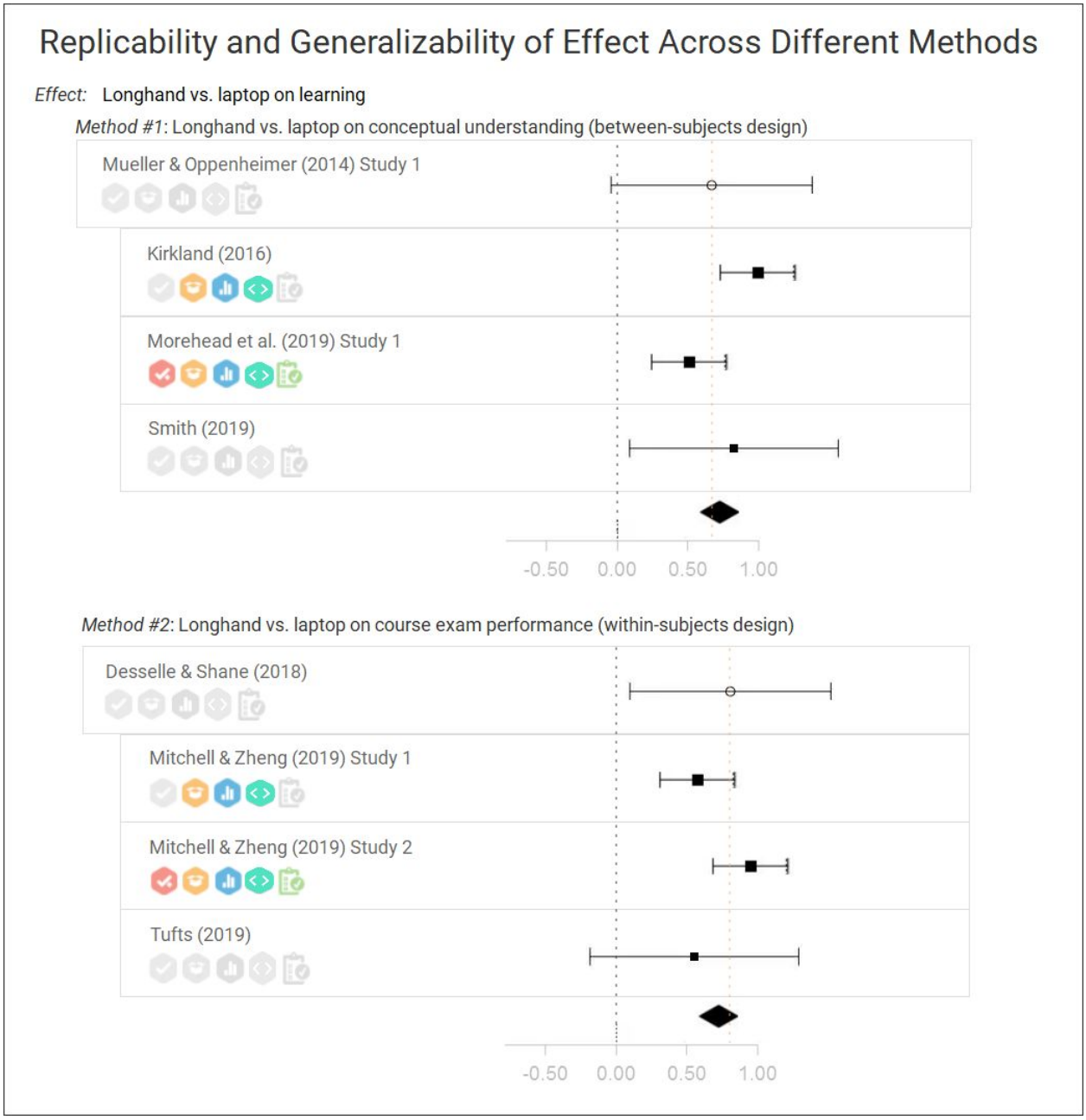


Figure 6. Idealized case whereby an effect (e.g., *longhand benefit effect*) is replicable and generalizable across two distinct methods.

The goal of the project is to implement a set of curation web-tools. These tools will be developed and integrated into the existing CurateScience.org platform, which currently exists as an early beta with the following very limited functionalities:

- Highly limited static homepage featuring example transparent articles.
- Very limited author page whereby researchers can label the transparency and replication information of their own articles (article-level curation only, rather than deeper study-level curation).

The new tools will substantially extend these basic functionalities, offering a fully featured product that will provide education researchers a suite of curation and compliance tools to verify the transparency, result reproducibility, replicability, and generalizability of education findings. For each of these tools, we will involve the community of education researchers in the design, testing, and improvement of the tools. We will solicit feedback on prototypes of the tools from education researchers who have conducted replications. Education researchers will be invited to be beta testers to test initial versions of the tools, which will then be improved based on these researchers' feedback.

Transparency Curation and Compliance Tools

Education researchers will be able to curate the transparency compliance of their own and other researchers' findings using specific tools available on the platform's homepage, search results page, article page, content tracking page, and replication collection page.

Homepage. The homepage will be substantially enhanced in the following ways:

- Improved search, filtering, and sorting of articles to better identify transparency-compliant findings (to replicate, generalize, meta-analyze, or teach about).

- Live article feed tailored to a user's research area and research interests (see Figure 7).
- Full-screen mode to efficiently view embeddable content in a side panel (see Figure 7).

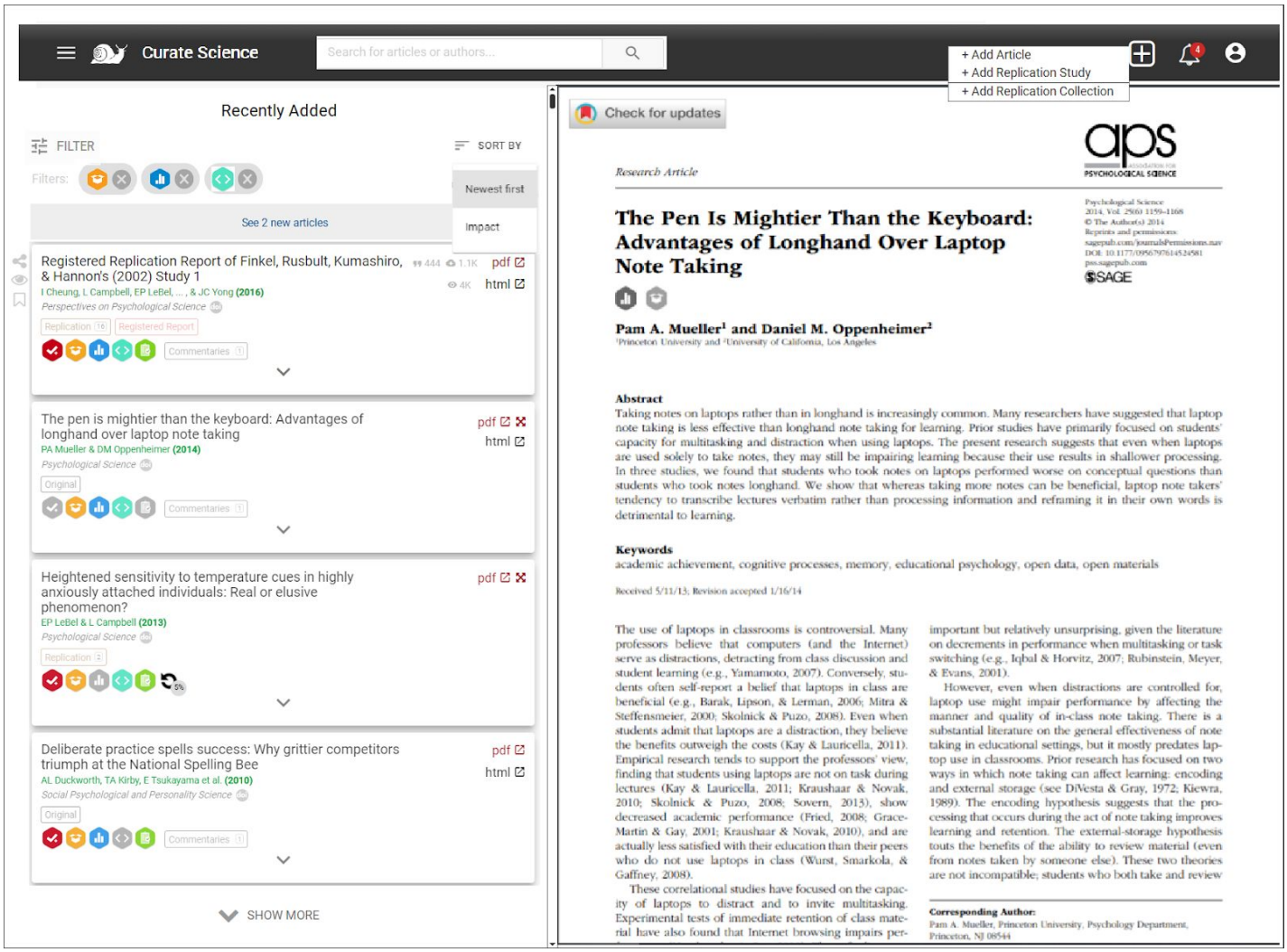


Figure 7. Prototype of improved homepage to identify and track transparency-compliant findings/articles.

Article page. The article page will show primary and secondary transparency information and include the following features (see Figure 8 below for prototype):

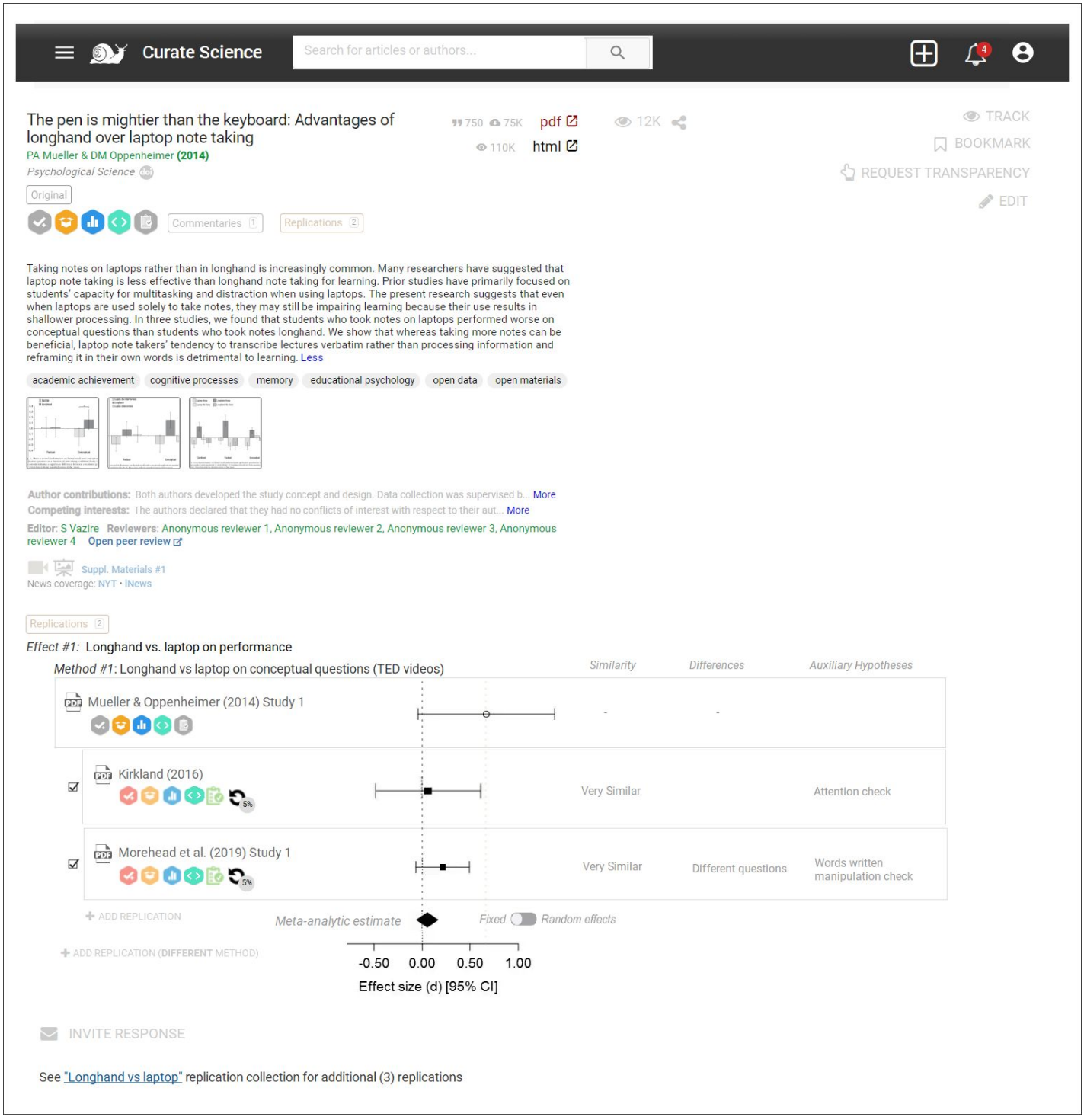


Figure 8. Prototype of article page, displaying (interactable) transparency and replication details.

- Study-level transparency curation.

- Request transparency of an article (e.g., open data, reporting standards compliance).
- Track articles to be automatically notified when new transparency, replication information, or follow-up commentaries become available.
- Upload files to the OpenScienceFramework.org (OSF) directly from an article card/page on Curate Science (via OSF's API).

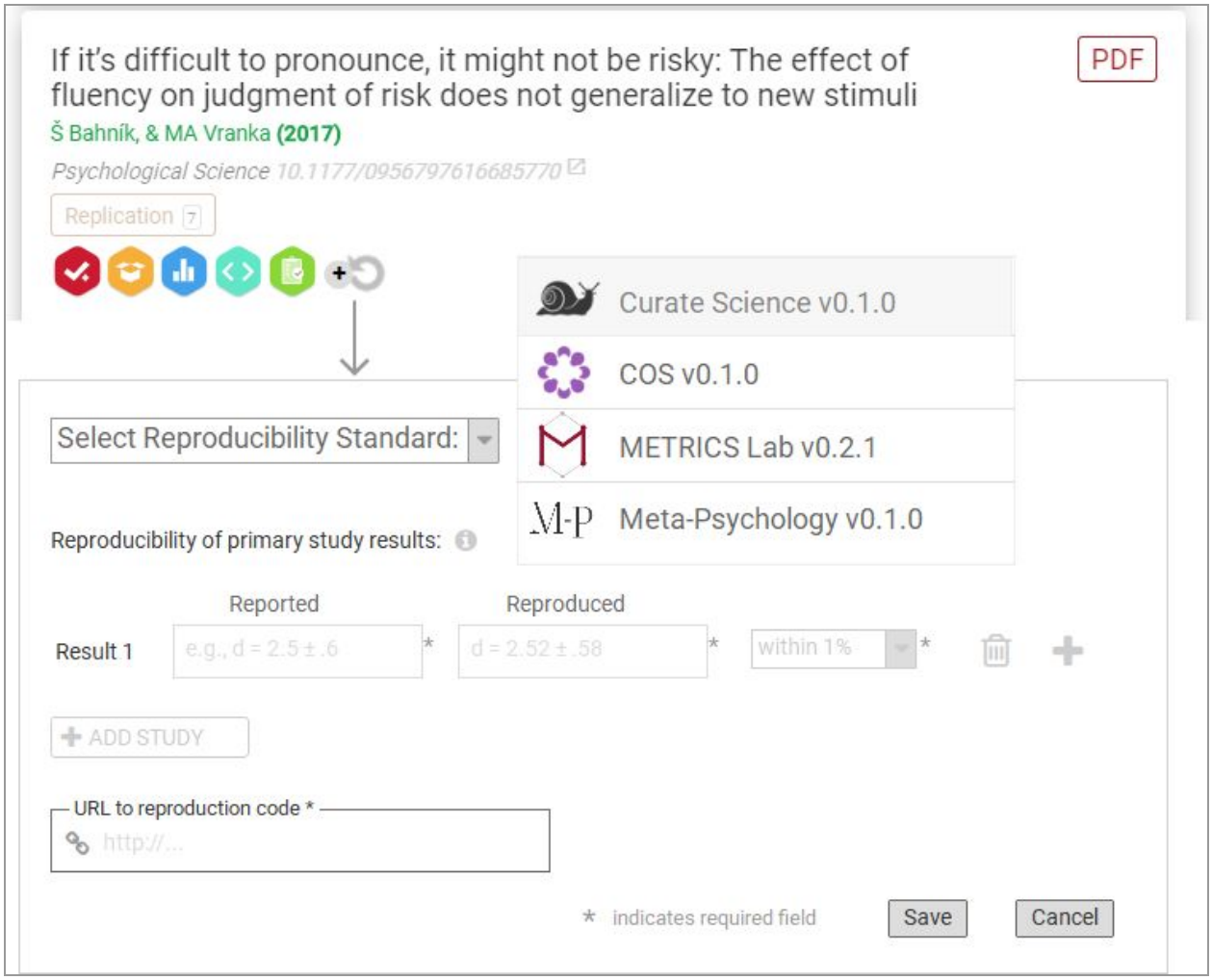


Figure 9. Prototype of form to save/post an article's result reproducibility information.

Result Reproducibility Curation and Compliance Tools

Result reproducibility curation tool. This curation tool will allow an independent researcher to save information regarding the results of an independent reproducibility check of a published paper's main results. A user will be able to post the reproduced result (compared to the reported result) for each primary outcome that constitutes the "primary substantive" finding of a study (e.g., Result 1, Result 2, and Result 3; for each study reported in an article; see Figure 9 above).

Result reproducibility compliance tool. The independent researcher will then be able to indicate an article's reproducibility status according to a specific reproducibility standard. For example, if using our proposed reproducibility standard (see Figure 9), the analyst will be able to indicate that an article's primary results are (overall) reproducible within 5% of reported results according to this standard (or according to an alternative reproducibility standard that may be more stringent, e.g., the journal *Meta-Psychology's* v0.1.0 reproducibility standard). Reproducibility results and compliance information will be cleanly displayed in a popup window (on hover) as is depicted in the prototype below (see Figure 10).

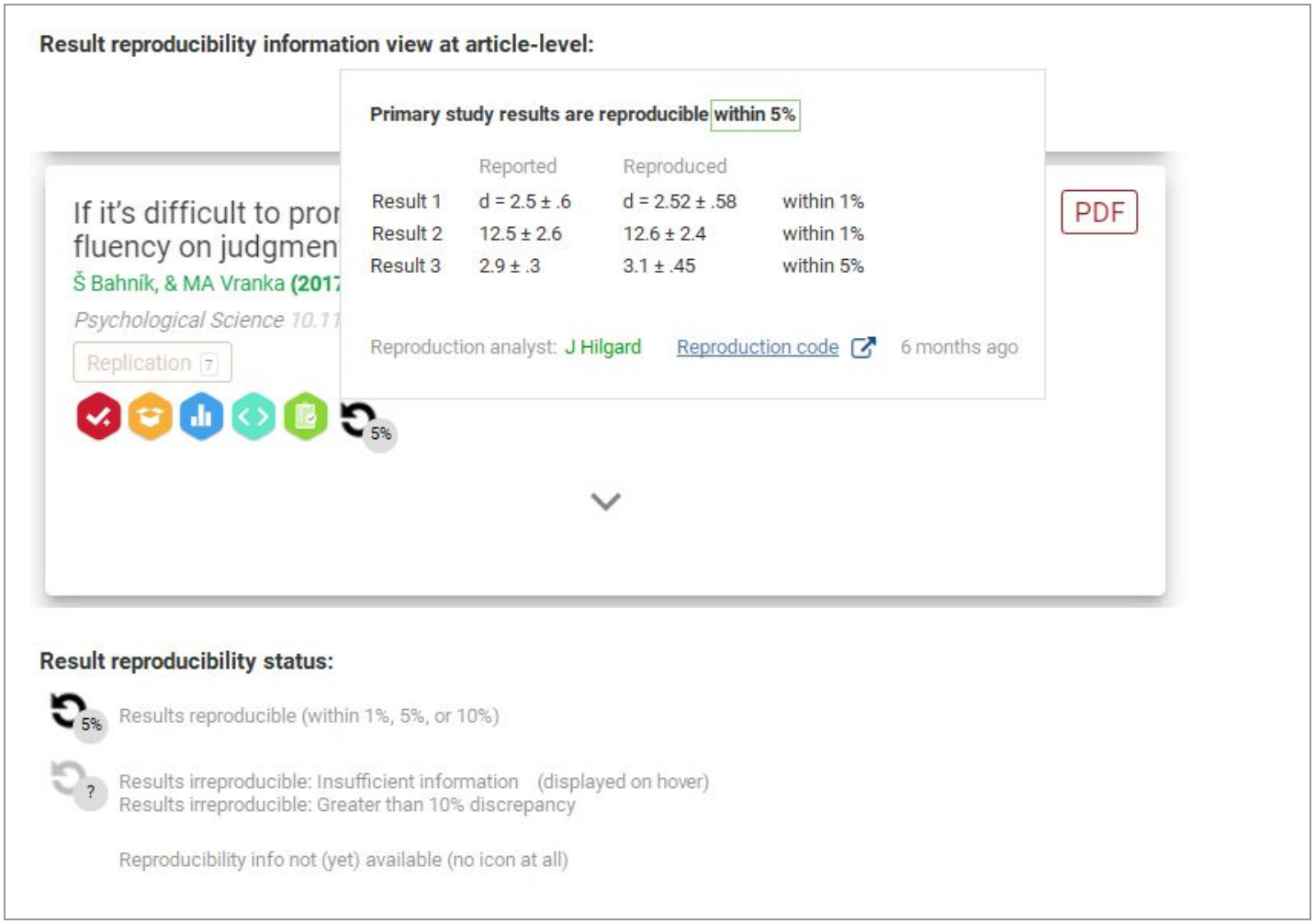


Figure 10. Prototype of how result reproducibility information will be displayed within an article card.

Replication Curation/Tracker and Replicability/Generalizability Compliance Tools

Replication curation/tracker tool. This will allow researchers to curate basic replication study information and link a replication to its original study parent. This includes an automatic method similarity categorization functionality, which ensures that a follow-up study is sufficiently methodologically similar to a previous study (LeBel et al., 2017; see Figure 11).

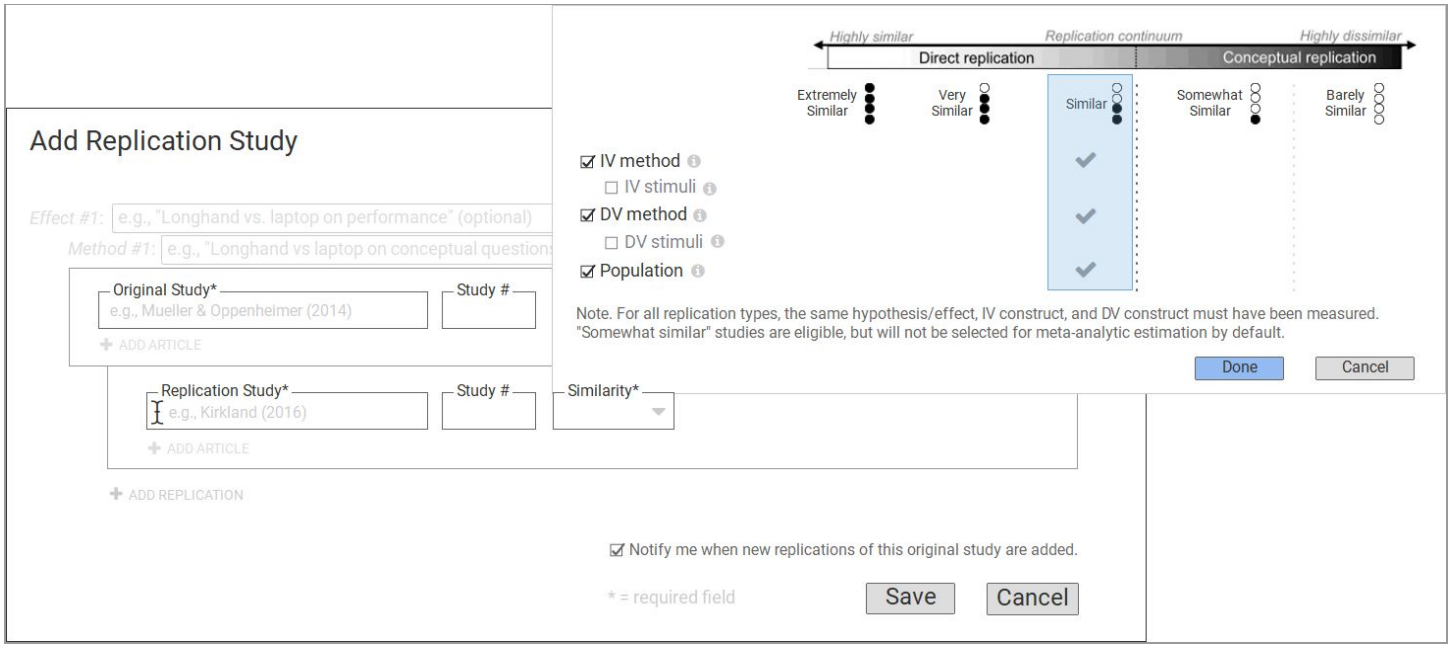


Figure 11. Prototype of method similarity categorization tool within the Add Replication Study window.

Replicability/generalizability compliance tool. This will allow researchers to organize replications of effects by general methods/populations to gauge the replicability and generalizability of an effect.[3] Replications will be organized in replication collections, which will allow researchers to meta-analyze *subsets* of replications based on their transparency/reproducibility compliance and other study characteristics (e.g., only include transparency-compliant replications in the meta-analytic estimate; see Figure 12 for prototype). As can be seen in Figure 12, the *longhand benefit effect* does not appear replicable using a between-subjects design, but does so via a within-subjects design.

[3] Replication collections can flexibly accommodate replications of effects across distinct methods originating from *more than one* original article.

Curate Science

Search for articles or authors...

New Collection

TRACK
BOOKMARK
EDIT

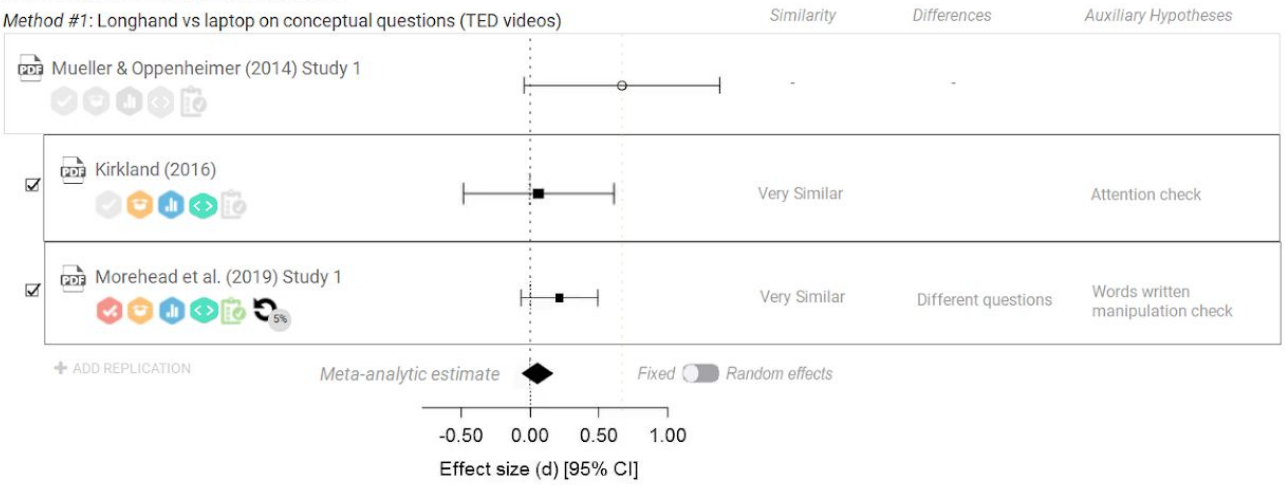
Longhand vs. laptop effects *

16K

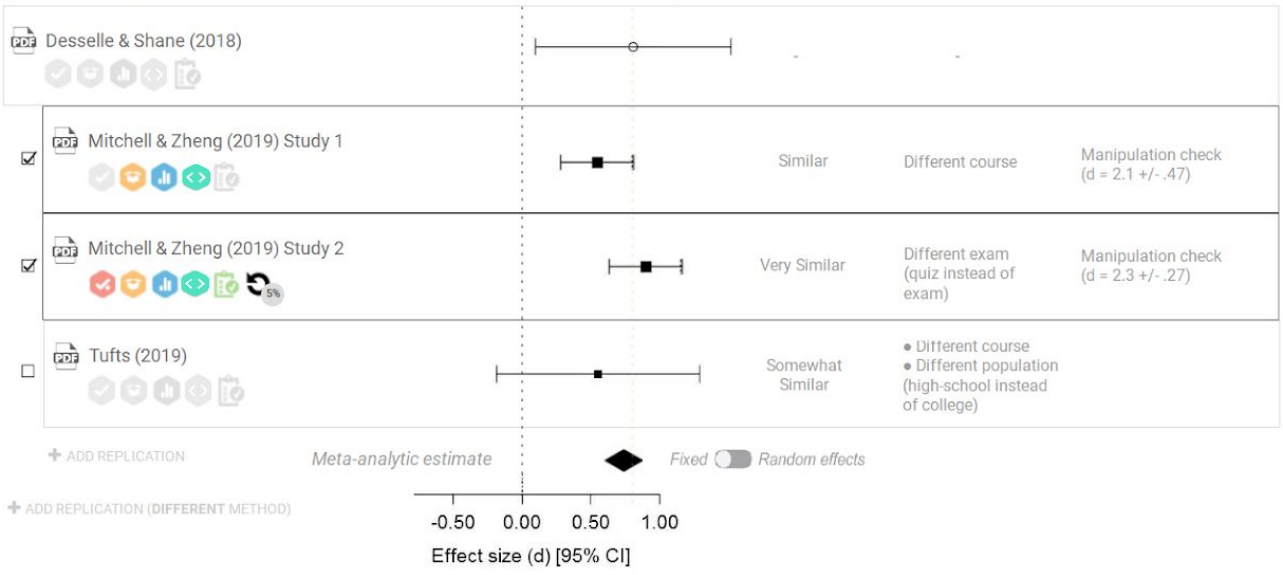
Replications of various longhand note taking vs. laptop on learning effects across different methods.

Effect #1: Longhand vs. laptop on performance

Method #1: Longhand vs laptop on conceptual questions (TED videos)



Method #2: Longhand vs. laptop on course exam performance (within-subjects design)



ADD REPLICATION (DIFFERENT EFFECT)

Original authors' response:

We applaud Balzarini et al.'s carefully conducted close replications of our previously reported playboy effect. We believe the effect might be sensitive to cultural and/or personality factors, which may be worth investigating in future research.

douglas.kenrids *November 3, 2018*

INVITE ADDITION OF REPLICATIONS

* = required field

Save | Save & continue later | Cancel

Figure 12. Prototype of replication collection that will allow (automatically-updating) meta-analysis of replications across distinct methods of observing an effect (see live demo).

To be fair to original authors (and ensure accountability of curators), authors will be able to respond/comment about replication attempts of their work (see Figure 12, bottom of the figure). Original authors will be automatically notified when new replications of their work are posted.

Strategies and Tools to Increase Adoption

The Curate Science platform is specifically designed to be selfishly beneficial to researchers, even if the platform is not widely adopted by the community. In particular, by maintaining their author page, researchers can easily organize, share, and disseminate open access links to their publications, public materials/data/code, and key figures/interactive charts (see early beta author page example).[4] However, the platform will be substantially more valuable with respect to accelerating scientific progress if it is widely adopted. Hence, we will design and implement the following specific tools to help accelerate broad community adoption:

- *Crowdsourcing functionalities*: This will allow any registered (non-author) researcher (e.g., graduate student) to curate an article's transparency/reproducibility information and add missing replications. To incentivize the quantity and quality of contributions, we will feature a user contribution leaderboard and user page that lists a user's recent edits/additions and contribution statistics (see prototype below, Figure 13).

[4] Maintaining an author page will also boost the impact of one's research and publicly signal one's commitment to scientific transparency and accountability.

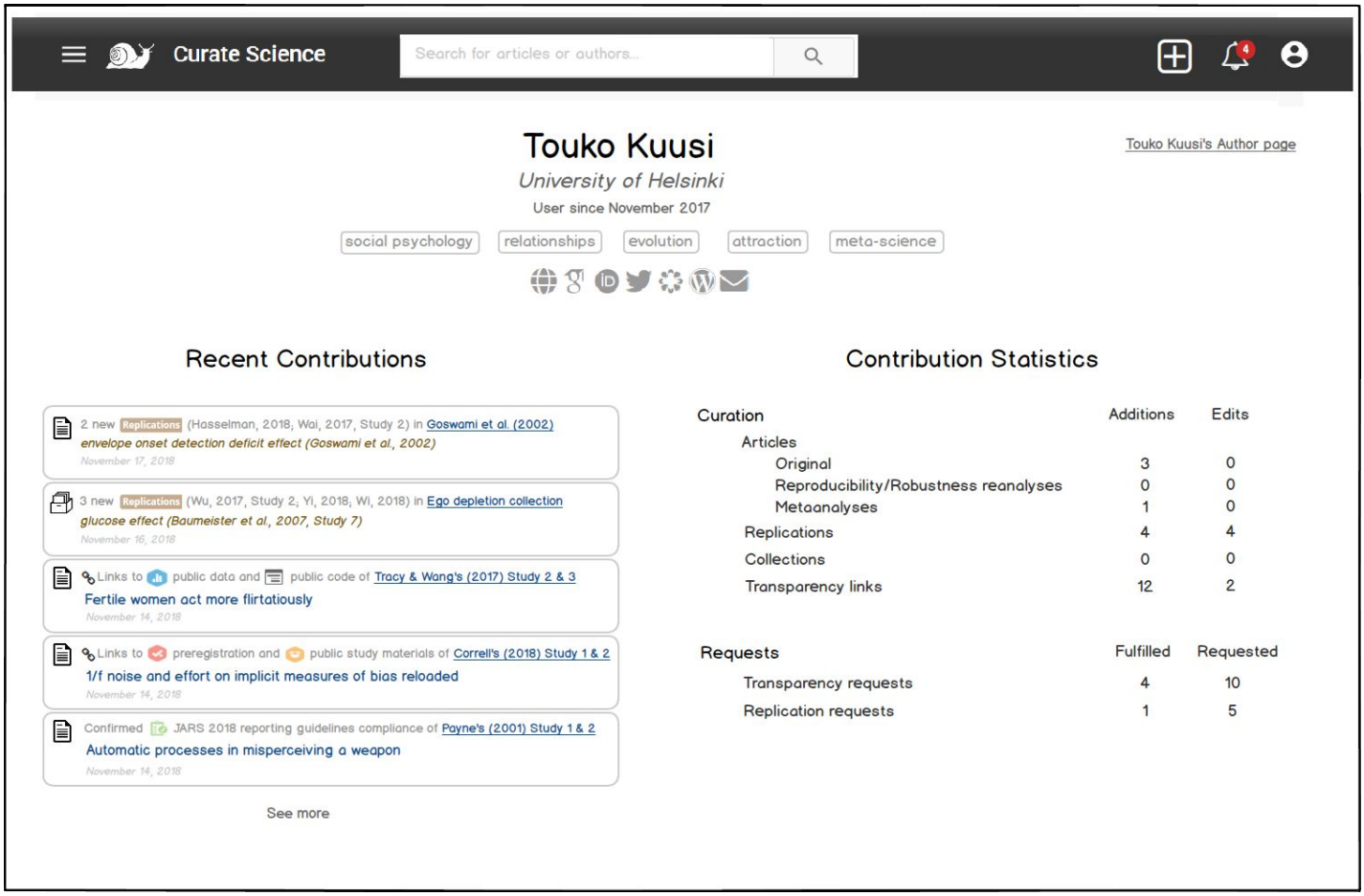


Figure 13. Prototype of user page that displays a user's contributions, which incentivizes curation contributions and makes users publicly accountable.

- *Transparency widgets for journals*: We will develop embeddable widgets that journals can use (within their submission portal) to ensure transparency compliance of submitted articles and an article list widget to display articles with labeled (and interactable) transparency/replication information.
    - *Transparency curation submission portal widget*: This will allow education journal editors to ensure all submitted articles comply with the appropriate transparency standards of their choosing at submission or upon manuscript acceptance (see Figure 14 for prototype).

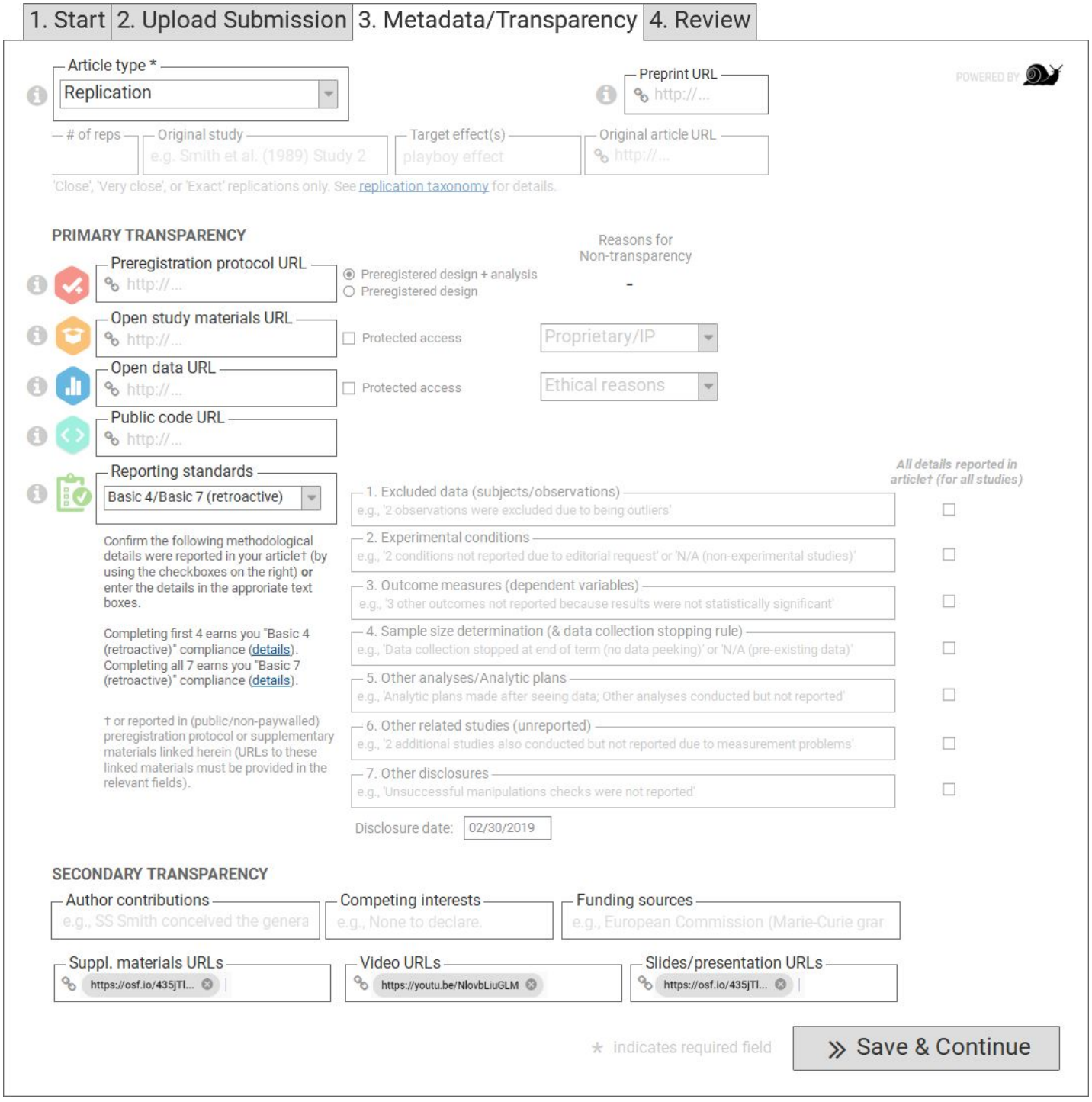


Figure 14. Prototype of transparency-compliance widget journals will be able to use within their submission portal.

- *Transparency compliance embeddable article list widget*: This will allow any (education) journal to display their issue article lists using Curate Science's transparency labeling and compliance system (see Figure 15 for prototype). This will make journals scientifically accountable given that any follow-up critical commentaries and/or follow-up replication studies added on Curate Science (or anywhere else) will be automatically linked and displayed within a journal's issue article lists.

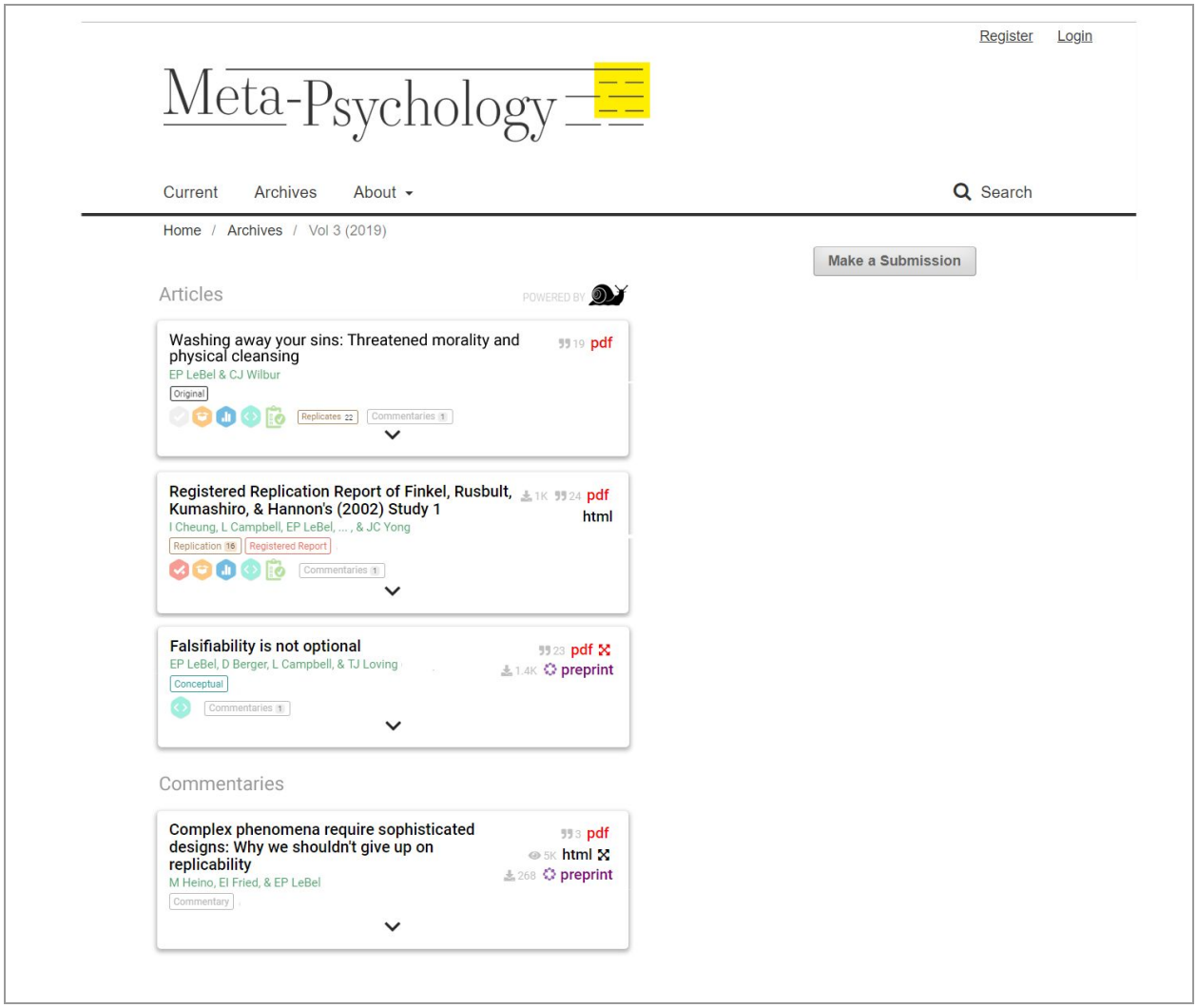


Figure 15. Prototype of article list widget journals will be able to use to display issue article lists of (automatically-updating) transparency-labeled articles.

- *Integration with related apps/initiatives*. Our system is designed using a modern open API architecture, which means its data will be seamlessly integratable with other transparency/replication-related websites and applications (e.g., our article transparency/replication information could be integrated/displayed within the article reference manager Zotero, as recently requested).

To further increase uptake, we will work with teachers of undergraduate and graduate methods courses to coordinate having their students curate transparency/reproducibility and replications as part of their course assignments (as has been offered by transparency maverick Uri Simonsohn).

Summary and Conclusion

Education researchers cannot currently determine the actual efficacy of interventions in producing desired learning outcomes because no system exists to check whether (education) studies meet basic transparency and replication standards. The curation system/tools that will be developed will address this problem by allowing (education) researchers to verify the transparency and replicability/generalizability of published education findings. This will generate an immense amount of value for education and other research stakeholders (see Figure 16).

	Value created/Benefits provided	
Stakeholder	**Immediate/Short-term**	**Long-term**
(Education) Researchers	• Identify transparency-compliant studies to replicate, extend/generalize, or meta-analyze • Author page: Curate open access links to publications, open materials/data/code, and interactive charts • Rich metadata for meta-science research (e.g., to understand predictors of replicability/generalizability)	• Accelerate self-correction of published literature • Accelerate development of educational theory and production of solid knowledge structure
Teachers/ Students	• Use platform to teach about transparency and replications (e.g., research methods courses) • Identify transparency- and replication-compliant findings/articles to teach about in substantive courses	• Identify and deploy credible educational techniques • Enhance learning/educational outcomes
Government /Nonprofit institutes	• Identify transparency-compliant studies for What Works Clearinghouse (education) and Cochrane Reviews (medical)	• Identify and deploy trustworthy educational or health interventions
Journals	• Use compliance system in submission portal to ensure submitted articles are transparency-compliant • Use article list issue widget to display transparency-compliant articles	• Enhance accountability of journals: Issue article lists/cards automatically update based on follow-up replications, commentaries, etc.
Universities	• Use compliance system to ensure professors' research is transparency-compliant and accountable to follow-up scrutiny/replications	• Accelerate scientific progress/cumulative knowledge development and applied innovations
Funders	• Use compliance system to ensure grantees' research is transparency-compliant and accountable to follow-up scrutiny/replications	• Accelerate development of applied solutions to solve society's most pressing (medical, social) problems
Public/ Journalists	• Use platform to look up the transparency/replication compliance of scientific findings discussed in the media	• Track credibility of evidence of personally relevant effects (e.g., dyslexia treatment; sugar intake)

Figure 16. Immediate and long-term value that will be created by the proposed transparency compliance system for various research stakeholders.

The time is ripe to develop such a transparency/replication curation system given that a strong demand already exists for it amidst the current "credibility revolution". Indeed, the research community regularly expresses on social media its desire and need for a more comprehensive/scaled up version of CurateScience.org's platform ([1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28]).

References

Bahník, Š., & Vranka, M. A. (2017a). Growth mindset is not associated with scholastic aptitude in a large sample of university applicants. *Personality and Individual Differences*, *117*, 139-143.

Bahník, Š., & Vranka, M. A. (2017b). General and domain-specific grit both show a negative association with scholastic aptitude in a large sample of university applicants. Retrieved from https://osf.io/nh88v/

Finnigan, K. M., & Corker, K. S. (2016). Do performance avoidance goals moderate the effect of different types of stereotype threat on women's math performance?. *Journal of Research in Personality*, *63*, 36-43.

Hanselman, P., Rozek, C. S., Grigg, J., & Borman, G. D. (2017). New evidence on self-affirmation effects and theorized sources of heterogeneity from large-scale replications. *Journal of Educational Psychology*, *109*(3), 405.

Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C., Kidwell, M. C., ... & Lenne, R. L. (2018). Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal Cognition. *Royal Society open science*, *5*(8), 180448.

Hedges, L. V., & Schauer, J. M. (2019). More Than One Replication Study Is Needed for Unambiguous Tests of Replication. *Journal of Educational and Behavioral Statistics*, 1076998619852953.

Inzlicht, M. & Jensen, L. E. (2014). Stereotype Spillover Effects on Women in Mathematics. *2014 NCUR*. Retrieved from

http://ncurproceedings.org/ojs/index.php/NCUR2014/article/download/1015/536

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.

LeBel, E. P., Berger, D., Campbell, L., & Loving, T. J. (2017). Falsifiability is not optional. Journal of Personality and Social Psychology, 113, 254-261. https://etiennelebel.com/documents/lbcl(2017,jpsp).pdf

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science*, *1*(3), 389-402. https://etiennelebel.com/documents/lebeletal(2018,ampss)a-unified-framework-to-quantify-the-credibility-of-scientific-findings.pdf

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), Criticism and the growth of knowledge (pp. 91–196). Cambridge, England: Cambridge University Press.

Li, Y., & Bates, T. C. (2017). Does mindset affect children's ability, school achievement, or response to challenge? Three failures to replicate. *Unpublished manuscript*. Retrieved from http://mrbartonmaths.com/resourcesnew/8.%20Research/Mindset/Mindset%20replication.pdf

Makel, M. C., & Plucker, J. A. (2014a). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, *43*(6), 304-316.

Makel, M. C., & Plucker, J. A. (2014b). Creativity is more than novelty: Reconsidering replication as a creativity act. Psychology of Aesthetics, Creativity, and the Arts, 8(1), 27–29.

Mehr, S. A., Schachner, A., Katz, R. C., & Spelke, E. S. (2013). Two randomized trials provide no consistent evidence for nonmusical cognitive benefits of brief preschool music enrichment. *PloS one*, *8*(12), e82007.

Mueller, P. A., & Oppenheimer, D. M. (2014). The pen is mightier than the keyboard: Advantages of longhand over laptop note taking. *Psychological science*, *25*(6), 1159-1168.

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Du Sert, N. P., ... & Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature human behaviour*, *1*(1), 0021.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... & Contestabile, M. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422-1425.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.

Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence?. *Perspectives on Psychological Science*, *7*(6), 528-530.

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence. *Psychological science in the public interest*, *9*(3), 105-119.

Popper, K. R. (1959). The logic of scientific discovery. London, England: Hutchinson.

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.

Stoet, G., & Geary, D. C. (2012). Can stereotype threat explain the gender gap in mathematics performance and achievement?. *Review of General Psychology*, *16*(1), 93-102.

Sunny, C. E., Taasoobshirazi, G., Clark, L., & Marchand, G. (2017). Stereotype threat and gender differences in chemistry. *Instructional Science*, *45*(2), 157-175.