Should We Communicate the Uncertainty of Educational Effects to Teachers?

Hugues Lortie-Forgues[1], Ut Na Sio[2], and Matthew Inglis[1]

[1]Loughborough University, Loughborough, UK

[2]University of Sheffield, Sheffield, UK

Corresponding author: Hugues Lortie-Forgues, Centre for Mathematical Cognition, Loughborough University, Loughborough, LE11 3TU. United Kingdom.

Email: h.lortie-forgues@lboro.ac.uk

Preregistered analysis plan, survey questionnaires, analysis code, data, and supplementary material are available at: https://osf.io/35nwp/

Funding details

This work was partially supported by Research England, via an Expanding Excellence in England grant to the Centre for Mathematical Cognition, and the Economic and Social Research Council [grant number ES/W002914/1].

Disclosure statement.

The author declares that there are no conflicts of interest related to this paper.

Abstract

There is a growing effort to inform teachers about the effects of various educational practices; however, the uncertainty associated with these effects is often very large and rarely reported. As a result, teachers may be unaware that the actual effects of interventions can often deviate substantially from what is reported. Should we communicate the uncertainty of educational effects to teachers? Doing so could support their decision-making, yet research from other fields (e.g., in risk communication) suggests that reporting uncertainty can be perceived negatively. We conducted two pre-registered studies to examine this question. Reporting uncertainty did not impact the perceived understandability of the information presented but decreased the perceived effectiveness of an intervention, suggesting that reporting uncertainty may influence teachers' decisions to implement an educational practice in their classrooms. However, we also found that reporting uncertainty leads to a decrease in trust in those who communicated the information, a particularly worrying observation, given the ubiquity and magnitude of uncertainty in educational research. We discuss the implications of these findings for research communicators, research evaluators, and initial teacher education.

*Keywords***:** Teachers, Communication, Uncertainty

Should We Communicate the Uncertainty of Educational Effects to Teachers?

Introduction

In the last 20 years, there has been a growing effort to inform teachers about the best educational practices. One approach has been to provide teachers with quantitative estimates of the effect that these practices have on student achievement, allowing teachers to compare the effectiveness of different interventions, weigh their impact against their costs, and ultimately select the most promising course of action. Although criticized (e.g., Simpson, 2017), this approach remains very popular. Books summarizing the effects of various interventions, such as Hattie's *Visible Learning for Teachers* (2012) have become widely popular (cited more than 7400 times at the time of writing, 3200 times since 2020). Online platforms, such as the What Works Clearinghouse (WWC) and the Education Endowment Foundation (EEF), which provide impact estimates for a variety of educational practices, also enjoy high popularity. In the UK, more than 70% of school leaders report using material from the EEF to inform their decisions (Higgins, 2022). Related efforts have been made to facilitate teachers' interpretation of these effects. Intuitive metrics, such as percentile gains and additional months of progress, have been developed to more clearly convey the effects of interventions to practitioners (e.g., Lipsey, 2012), and widely used benchmarks have been established for comparing the impact of interventions to typical effects observed in the field (e.g., Kraft, 2020).

Teachers care about an intervention's impact on student achievement when deciding whether to implement it (Holland et al., 2019), and, as such, it is no surprise that they refer to this information when making decisions. Importantly, however, it is becoming increasingly clear that the *uncertainty* associated with impact estimates of educational interventions is problematically large (e.g., Lortie-Forgues & Inglis, 2019). Intervention effects will often deviate substantially from the impact estimate reported, meaning that ignoring uncertainty could lead teachers to greatly misjudge the effect an intervention may have on their students, or, for example, to conclude that two interventions with different point estimates will produce different effects when in fact they may not.

While there appears to be a need for teachers to consider uncertainty, and a need for research communicators to clearly communicate this information, evidence from other fields (e.g., risk communication) suggests that information about uncertainty is sometimes ignored, and can sometimes be perceived negatively, potentially diminishing teachers' engagement with research (e.g., Johnson & Slovic, 1995; van der Bles et al., 2020). Unfortunately, our understanding of the way teachers respond to uncertainty about intervention effectiveness is limited. Do teachers perceive and engage with information about uncertainty? Could communicating uncertainty reduce teachers' engagement with the information? To our knowledge, no studies have yet examined these questions, which have important implications for how we communicate research outcomes. Moreover, if teachers ignore uncertainty, or perceive it negatively, these questions also have implications for initial teacher education. We conducted two studies to address this knowledge gap.

Uncertainty of Educational Effects

Evaluations of educational interventions are prone to high levels of uncertainty. To generate a realistic estimate of an intervention's impact, evaluation studies must closely mimic the settings where the interventions are meant to be implemented and follow rigorous procedures. This often implies that interventions will be tested in uncontrolled environments such as classrooms, implemented by multiple teachers, administered to heterogeneous samples of students, and that randomization to conditions will occur at the school level to prevent contamination—factors that compound the uncertainty of the findings.

Unsurprisingly, even the largest and most rigorous evaluation studies often produce highly uncertain findings. Consider the sampling uncertainty of large-scale educational randomized controlled trials (RCTs), such as those commissioned by the EEF or the National Center for Education Evaluation and Regional Assistance (NCEE). These trials are large (median number of participants > 2000), yet their confidence intervals (CI) tend to be very wide. As noted by Lortie-Forgues and Inglis (2019), the median width of the 95% CI in the RCTs conducted by the EEF and NCEE is 0.24 SDs (i.e., ± 0.12 SDs) — a large interval considering the small effect sizes typically observed in these trials (i.e., 80% of effect sizes range between -0.06 SDs and 0.20 SDs). As a result, (1) actual outcomes of interventions are likely to depart substantially from the point estimate communicated, (2) many trials are inconclusive, with many confidence intervals including zero, and (3) comparing interventions based on the point estimates reported is rarely meaningful due to the extensive overlap of most of the confidence intervals. Note that the sampling uncertainty discussed above is just one source of uncertainty, and that the overall uncertainty of educational effects, which includes other sources of uncertainty (e.g., methodological, analytical), is likely to be greater.

In addition to being large, the uncertainty of educational effects also *varies* between studies. For example, while the CI widths associated with effects found in EEF and NCEE trials are around 0.24SDs, some are noticeably narrower, with the smallest 10% falling below 0.12 SDs (i.e., ± 0.06 SDs). Precise estimates are usually obtained when the nature of the intervention allows for an unusually large sample or student-level randomization. In these studies, the real effects are more likely to align with the reported effects. By contrast, some effects are surrounded by very wide CIs — with the widest 10% of the intervals exceeding 0.46 SDs (i.e., ± 0.23 SDs) — suggesting that the point estimate provides little information about the potential intervention. The fact that uncertainty in rigorous large-scale trials is both large and variable suggests that it should be carefully considered when making decisions using the findings of those trials.

Teachers and Uncertainty

Teachers are likely to be particularly attuned to the magnitude and variation in uncertainty of educational effects and to use this information in their decision-making. Given the limited time they have to improve their students' achievement, and the potential consequences of not doing so, teachers' decisions are likely oriented toward minimizing potential negative outcomes (see Studdert et al., 2005, for related decision patterns in health professionals). Teachers may, for example, be wary of a large point estimate if the associated uncertainty suggests that the true effect could actually be low or negative, and may prioritize interventions with more certain effects. Failing to report uncertainty transparently could prevent teachers from adopting this cautious approach. Teachers may also assume that uncertainty is negligible, which could lead to frustration if outcomes are less favorable than expected.

While reporting uncertainty is likely to benefit teachers' decision making, the way teachers respond to being presented with uncertainty remains an open question. Research from other domains (e.g., risk communication to the general public) shows that some people appear to overlook information about uncertainty entirely (e.g., Johnson & Slovic, 1995). When considered, uncertainty can also be perceived negatively, potentially reducing engagement with the information presented. For example, uncertainty can be perceived as confusing (Schapira et al., 2001; Peters et al., 2007), and can reduce trust, both in the information presented, and in the people communicating that information (van der Bles et al., 2020). Both ease of understanding and trust are important influences on the extent to which people engage with a source of information (e.g., Davis, 1989; Kelton et al., 2008).

While these findings have implications for reporting uncertainty, it remains unclear whether they will be observed in teachers. There is evidence that teachers differ from the general population in potentially relevant ways, such as showing higher levels of risk aversion (Ayaita & Stürmer, 2020) and mathematics anxiety (Hembree, 1990). Moreover, the different contexts in which previous research on the communication of uncertainty has been conducted raise questions as to whether the same findings will apply when communicating the effects of educational interventions. Given the potential impact on teachers' engagement with research, it is important to investigate whether these patterns hold true for teachers presented with educational effects. The present studies assess the impact of reporting uncertainty on four aspects of teachers' perceptions:

Perceived Uncertainty. We aim to explore whether reporting uncertainty leads to changes in perceived uncertainty. If this is the case, it would suggest that teachers are indeed perceiving and processing the information presented.

Perceived Understandability. Presenting uncertainty, rather than simply a point estimate, introduces additional, potentially confusing, information. Given the relation between understandability and engagement (e.g., Davis, 1989), if reporting uncertainty decreases perceived understandability, research communicators should carefully consider the implications of reporting this information.

Perceived Trustworthiness. Trustworthiness is another important predictor of engagement—people are more likely to engage with information they perceive as trustworthy (Kelton et al., 2008). Although reporting uncertainty could potentially increase trust (indeed, as uncertainty is ubiquitous, reporting it shows a commitment to transparency), research shows that it can also reduce trust (e.g., Johnson & Slovic, 1995; van der Bles et al., 2020). If this effect is observed among teachers, research communicators should, again, weigh the benefits and costs of communicating this information. Trust can encompass different dimensions: Study 1 focuses on trustworthiness of the information presented, while Study 2 examines trust in (i) the estimate of impact and (ii) the people who communicated the information.

Perceived Effectiveness. Observing a change in the perceived effectiveness of an intervention when reporting uncertainty would have important implications for research communicators. Given teachers' interest in the impact of educational interventions when deciding on whether to implement a given intervention (Holland et al., 2019), this would be a compelling case for research communicators to report uncertainty — not doing so would deny teachers information they would likely use in their decision-making. Also of interest is the direction of the change induced by uncertainty: it could increase or decrease the perceived effectiveness of an intervention.

Ways to Communicate Uncertainty

In this study, we opted to present uncertainty as a numerical range (see Table 1), a common approach recommended by numerous reporting guidelines (e.g., CONSORT statement [Schulz et al., 2010], Cochrane handbook [Higgins, et al., 2023], GRADE working group [Hultcrantz, et al., 2017], American Psychological Association [APA, 2020]). There are other ways to convey uncertainty. For example, the What Works Clearinghouse (WWC) and the Education Endowment Foundation (EEF) both use tiered systems to qualify the level of certainty associated with a finding. The WWC's tiers (i.e., ESSA tiers of evidence) range from 1 (highest) to 3 (lowest), whereas the EEF uses a padlock rating from 1 (lowest) to 5 (highest) — both determined by criteria such as type of research design, sample size, and amount of attrition (EEF, 2019, WWC, 2022). We opted against using tiered categories to represent uncertainty in our study, as they do not provide clear information as to how readers should adjust their expectations and decision-making (van der Bles et al., 2019). For example, to be considered tier 1 (highest level of certainty), studies in the WWC must have at least a total of 350 participants (WWC, 2022) — meaning an effect could be surrounded by a confidence interval exceeding 0.40 SDs (i.e., ± 0.20 SDs)[1], a high level of uncertainty given the typical size of educational effects (e.g., Kraft, 2020). The ambiguity of tier systems also means that, in the present study, different participants may interpret the same level of uncertainty very differently, potentially obscuring the effects we seek to examine. Using tiered categories would also limit the comparability of our findings with studies conducted in other fields, many of which report uncertainty using numerical ranges.

While numerical ranges have limitations — for example, not all sources of uncertainty are easily quantifiable (e.g., attrition) and can be incorporated into a numerical range — they do provide a

---

[1] This was calculated using the formula from Borenstein et al. (2021): $d \pm 1.96\sqrt{\frac{n_1+n_2}{n_1 n_2} + \frac{d^2}{2(n_1+n_2)}}$ with $n_1 = n_2 = 175$ (total n = 350). The width of the confidence interval will be similar for a range of plausible values of d.

transparent means of communicating uncertainty. Moreover, people often prefer numerical representations of uncertainty over less transparent alternatives (e.g., Erev & Cohen, 1990), and these tend to be interpreted more consistently across individuals (Budescu et al., 1988).

Study 1

The main goal of Study 1 was to examine how reporting an intervention effect with a numerical range of uncertainty influences teachers' perceptions, particularly the extent to which it affects: (1) the amount of uncertainty they perceive about the impact of an intervention (perceived uncertainty), (2) the extent to which they think the description of the impact is easy to understand (perceived understandability), (3) the degree to which they find the information presented trustworthy (perceived trustworthiness of the information), and (4) how effective they perceive the intervention is at improving student achievement (perceived effectiveness of the intervention). We examine how these vary across three different levels of uncertainty: no uncertainty mentioned, low uncertainty, and high uncertainty—the latter two levels chosen to be distinct, yet aligned with levels of sampling uncertainty typically observed in educational research.

Another goal was to examine whether the impact of uncertainty is influenced by the metric used to report an intervention's effect. In education, intervention effects are usually reported in units of standard deviation (SD) but are often translated into more intuitive metrics before being communicated to teachers. Previous research has shown that teachers have strong preferences about the effect size metrics, and that different effect size metrics induce very different perceptions of an intervention's effectiveness (Lortie-Forgues et al., 2021). In this study, we examined the potential interaction between level of uncertainty and metric (more specifically: additional months of progress, percentile gain, and raw test score).

Methods

Participants

Four hundred and ninety-five teachers (88% female; Mean age: 40.22 years; *SD*: 9.95) were recruited from social media (*N* = 255, Facebook via teacher-only private groups) and Prolific (*N* = 240) to take part in an online survey. They had a mean of 12.76 years of teaching experience (SD: 8.94), worked mainly in the South (50%), the North (24%), and the Middle (23%) of England, and were mainly teaching in primary (45%) or secondary (41%) schools. See Table S3 for a detailed description of the sample and evidence of baseline equivalence between conditions, and Table S5 for a comparison of our sample with census data[2].

Survey Attrition

A total of 129 participants started but did not complete the survey. They were excluded from our analysis, as per our preregistration plan. Of those, 105 completed the demographic questions at

[2] This project was approved by the Loughborough University Ethics Committee (Project ID: 4296). In both Study 1 and Study 2, participants were required to agree to an informed consent form at the beginning of the survey before proceeding.

the start of the survey, and 94 progressed as far as being randomized to one of the three experimental conditions (post-randomization attrition was 16%, i.e., 94/(94 + 495)). Non-completers were similar to those included in the study across various demographic variables, but they differed in terms of the topics they teach (being less likely to teach multiple topics) and, to a lesser extent, in the types of schools they teach at (with fewer from state and special schools; however, this latter difference was not significant after correcting for multiple comparisons).

Importantly, we found no evidence that the number, or the demographics, of non-completers differed between the three experimental conditions (i.e., no evidence of differential attrition; see Table S9). We will nevertheless examine the potential impact of attrition on our findings later in the Robustness Analyses section.

Study Design

The study followed a mixed design with the level of uncertainty (no uncertainty mentioned, low uncertainty, and high uncertainty) manipulated between subjects and the metric used to report the intervention impact (additional months of progress, percentile gain, and raw test score) manipulated within subjects. The impact of these manipulations was measured on four dependent variables: (1) perceived uncertainty, (2) perceived understandability, (3) perceived trustworthiness of the information presented, and (4) perceived effectiveness of the intervention.

Procedure

Participants were informed that they would be presented with a brief description of the effectiveness of three different educational interventions. They were told that all three interventions were aimed at improving the performance of typical students on the mathematics test of the Key Stage 2 national curriculum assessments (KS2 mathematics; a well-known standardized test among teachers in England, our target participants), that they were of the same duration and cost, and that their effectiveness had been determined in the same manner: by comparing the performance of a group receiving the intervention to the performance of a group receiving normal instruction (i.e., ‘business as usual’) on the KS2 mathematics test.

Having been presented with the instructions, participants were randomly allocated to one of the three uncertainty conditions (no uncertainty mentioned, low uncertainty, and high uncertainty). All participants were presented with three vignettes (vignette a, b, and c, presented one at a time), each describing the impact of a hypothetical educational intervention using a different effect size metric: (a) additional months of progress, (b) percentile gain, and (c) raw test score. The order of the vignettes was randomized within each participant, and the effect of the intervention presented — despite being reported in different metrics — all corresponded to an effect size of 0.15 SDs[3], although participants

---

[3] Conversion into months of progress was based on EEF’s months of additional progress measure guideline [EEF, 2023]. Percentile Gain was computed in accordance with the WWC Procedures Handbook (What Works Clearinghouse, 2022). Test Score translation was based on the 2019 KS2 Math test mean scaled score (105), and standard deviation (7.27) obtained via England's Department for Education (DFE, 2019). Due to Covid-19, this was the most recent data available at the time the study was conducted.

were not made aware of this. Table 1 shows vignette “a”, which describes the effect of the intervention in terms of additional months of progress.

In addition to these descriptive vignettes, participants in the low- and high-uncertainty conditions received an additional statement describing a range of uncertainty, the wording of which was modelled on past studies (e.g., Johnson & Slovic, 1995; Lipkus et al., 2001). For example, when the intervention impact was reported in additional months of progress, participants in the *low*-uncertainty condition received this additional statement: “However, the results showed this figure could be as high as 3 months, or as low as 1 month.” while participants in the *high*-uncertainty condition received this additional statement: “However, the results showed this figure could be as high as 4 months, or as low as 0 months”. Values for the range were matched across metrics[4].

The width of the range in the low and high uncertainty conditions corresponded to 0.16 SDs (i.e., ± 0.08 SDs) and 0.30 SDs (i.e., ± 0.15 SDs), respectively. The low uncertainty condition was chosen to reflect a level of uncertainty lower than what is typically observed in the field. For example, at the time of writing, only 18% of the effects from rigorous, large-scale RCTs commissioned by the EEF have a 95% CI narrower than the range reported in that condition. By contrast, the high uncertainty condition was chosen to represent a level that is high, but not unrealistic. At the time of writing, close to 40% (i.e., 39.2%) of the effects from rigorous, large-scale RCTs commissioned by the EEF have a 95% CI wider than 0.30 SDs. See Table S1 for the complete set of vignettes presented across all conditions.

After each vignette, all participants were presented with four questions measuring their: (1) perceived uncertainty, (2) perceived understandability, (3) perceived effectiveness of the intervention, and (4) perceived trustworthiness of the information. See Table S2 for questions.

In addition, teachers were asked to report their gender, age, years of teaching experience, teaching level, type of school, geographical location, teaching subject, as well as indicating their familiarity with the statistical concept of effect size, their familiarity with the KS2 Math test (i.e., the outcome measure in the vignettes), whether they had ever heard of the impact of an intervention being described in months of progress, their perceived mathematical ability, and their perceived level of risk aversion. Our analysis plan was preregistered prior to data collection[5] and can be inspected, along with the survey questionnaire, justification for sample size, analysis code, and data, at: https://osf.io/35nwp/.

[4] We assumed no additional uncertainty arises from the growth estimates used in the conversion into months of progress.
[5] We took considerably longer to recruit our target sample than the four months stipulated in our pre-registration (see details in Table S7).

Table 1

Vignettes Describing Interventions' Impact in Additional Months of Progress by Reported Uncertainty Level.

Level of Uncertainty	Vignette used in the study
Not Reported	Intervention A had an average impact of 2 additional months' progress. In other words, the pupils receiving the intervention made, on average, 2 months' more progress than the pupils not receiving the intervention.
Low	Intervention A had an average impact of 2 additional months' progress. In other words, the pupils receiving the intervention made, on average, 2 months' more progress than the pupils not receiving the intervention. However, the results showed this figure could be as high as 3 months, or as low as 1 month.
High	Intervention A had an average impact of 2 additional months' progress. In other words, the pupils receiving the intervention made, on average, 2 months' more progress than the pupils not receiving the intervention. However, the results showed this figure could be as high as 4 months, or as low as 0 months.

Note: For vignettes describing intervention impacts in terms of Percentile Gain and Raw Test Score, refer to the Supplementary Material Table S1.

Results

Figure 1 shows the mean ratings for the four dependent variables by uncertainty level, and Table 2 shows their correlations.

Preregistered Analyses

We preregistered four analyses (i.e., four mixed-design ANOVAs), one for each outcome variable[6]. We describe the results for each outcome in turn.

Perceived uncertainty. Reporting uncertainty influenced the perceived uncertainty about the impact of the intervention, as indicated by our main effect of level of uncertainty, $F(2,492) = 5.40$, $p = .005$, $\eta^2 = .02$. Post hoc comparisons showed that perceived uncertainty in both the high-uncertainty condition ($M = 4.08$, $SD = 1.01$) and the low-uncertainty condition ($M = 4.06$, $SD = 0.99$) was significantly higher than in the no-uncertainty condition ($M = 3.76$, $SD = 1.03$; $p = .01$, $d = 0.31$, and $p = .02$, $d = 0.30$, respectively).

We also observed a significant main effect of metric type, $F(1.96, 964.44) = 7.13$, $p < .001$, $\eta^2 = .01$. Effects reported in percentile gains ($M = 4.12$, $SD = 1.34$) were perceived as more uncertain

[6] We use the anova_test() function in the rstatix R package (Kassambara, 2023), which automatically applies the Greenhouse-Geisser sphericity correction to within-subject factors that violate the sphericity assumption. Redoing the analysis without this correction does not influence any of our conclusions.

than effects reported in additional months of progress (M = 3.84, SD = 1.43, p < .001, d = 0.20)[7], while effects reported in raw test scores (M = 3.94, SD = 1.48) did not significantly differ from either (vs percentile gain: p = 0.053, d = 0.13; vs months of progress: p = 0.58, d = 0.07). This pattern was not unexpected and was consistent with Lortie-Forgues et al.'s (2021) findings that intervention effects reported as percentile gains were perceived as less informative, less helpful, and less understandable, than effects reported as months of progress or as raw test scores. We found no evidence that the impact of reporting uncertainty varied between effect size metrics — i.e., there was no significant interaction between level of uncertainty and effect size metric, $F(3.92, 964.44) = 1.56$, $p = .18$, $\eta^2 = .006$.

Perceived understandability. Despite increasing the length of the description and quantity of numerical information presented, we found no evidence that reporting a range of uncertainty impacted teachers' perceptions of understandability, $F(2, 492) = 1.35$, $p = .26$, $\eta^2 = .005$. We found a large effect of metric type, $F(1.97, 968.41) = 61.81$, $p < .001$, $\eta^2 = .11$. Irrespective of the level of uncertainty reported, effects reported as percentile gains (M = 3.88, SD = 1.68) were perceived as harder to understand than those reported as raw test scores (M = 4.67, SD = 1.60, p < .001, d = 0.48) or months of progress (M = 4.83, SD = 1.64, p < .001, d = 0.57), replicating Lortie-Forgues et al.'s (2021) findings. Again, we found no significant interaction between level of uncertainty and metric type, $F(3.94, 968.41) = 0.427$, $p = .79$, $\eta^2 = .002$.

Perceived trustworthiness of the information. Reporting uncertainty influenced the perceived trustworthiness of the information presented, $F(2, 492) = 5.87$, $p = .003$, $\eta^2 = .02$. Information was perceived as being less trustworthy when uncertainty was high (M = 4.07, SD = 1.01) compared to when no uncertainty was mentioned (M = 4.43, SD = 0.94, p < .002, d = 0.37). Perceived trustworthiness in the low uncertainty condition (M = 4.27, SD = 1.01) fell in between the two conditions and did not differ significantly from either. We found no significant main effect of metric type, $F(1.94, 954.41) = 0.700$, $p = .49$, $\eta^2 = .001$, and no significant interaction between level of uncertainty and metric type $F(3.88, 954.41) = 0.75$, $p = .55$, $\eta^2 = .003$. As shown in Table 2, perceived trustworthiness of the information was also negatively associated with perceived uncertainty, as was the case in van der Bles et al.'s (2020) study (see their supplementary material Figure S4).

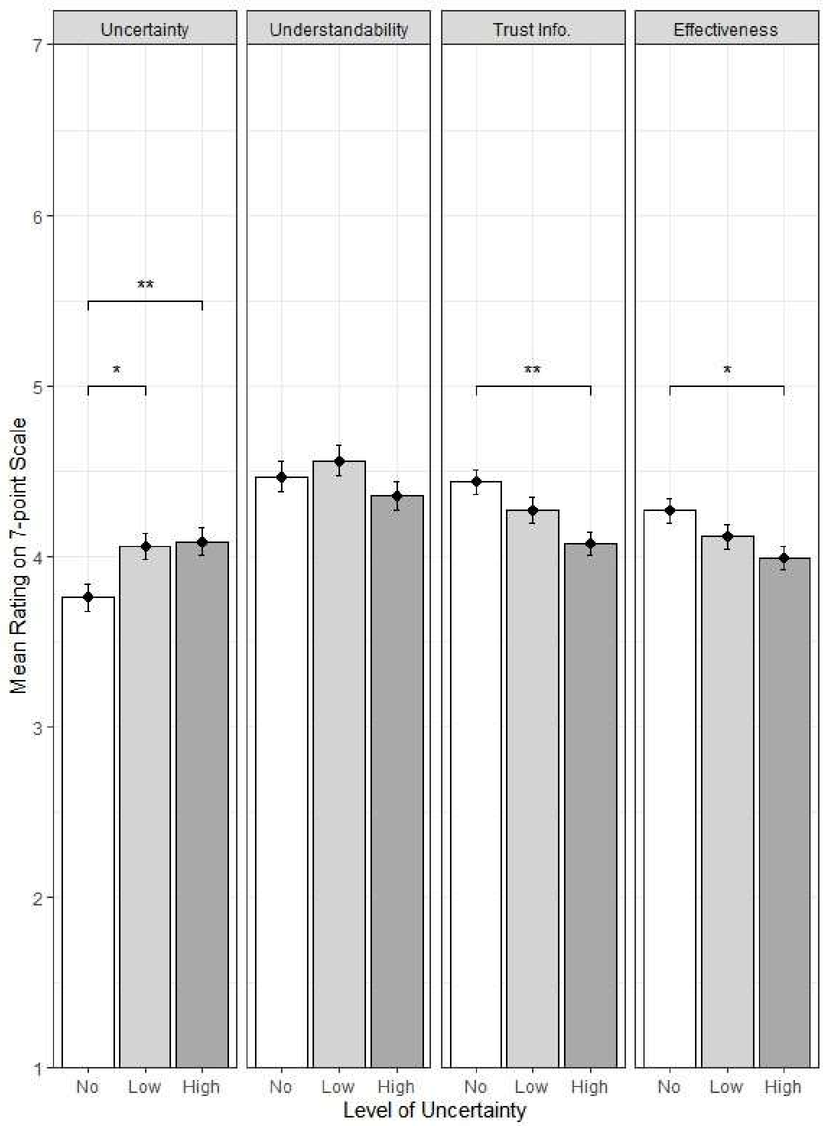
Perceived effectiveness of the intervention. Reporting uncertainty influenced the perceived effectiveness of an intervention, $F(2, 492) = 4.035$, $p = .018$, $\eta^2 = .02$. Interventions were perceived as less effective when uncertainty was high (M = 3.99, SD = 0.86) compared to when no uncertainty was mentioned (M = 4.27, SD = 0.92; p = .01, d = 0.31), while perceived effectiveness in the low uncertainty condition (M = 4.11, SD = 0.91) fell in between the two conditions and did not differ significantly from either. We also observed an effect of metric, $F(1.97 ,967.34) = 86.25$, $p < .001$, $\eta^2 =$

[7] Cohen's d for the effects of metric type (a within-subject factor) were computed by dividing the mean differences by the pooled standard deviation, ignoring the correlation between repeated measurements. This was done to allow comparison with the effect sizes for the different levels of uncertainty (a between-subject factor).

.15. Effects reported as raw test scores were perceived as being less effective (*M* = 3.57, *SD* = 1.45) than those reported as percentile gains (*M* = 4.22, *SD* = 1.29; $p < .001$, $d = 0.47$), which in turn was perceived as less effective than those reported as months of progress (*M* = 4.57, *SD* = 1.30; $p < .001$, $d = 0.27$), replicating the same ordering observed by Lortie-Forgues et al. (2021). We found no evidence of an interaction between level of uncertainty and metric type, $F(3.93, 967.34) = 0.71$, $p = .59$, $\eta^2 = .003$.

Figure 1

Mean Ratings to the Four Dependent Variables by Level of Uncertainty Reported



Note: The data presented here is aggregated across metric types. Error bars represent standard errors. Significance levels are indicated as follows: $*p < 0.05$, $**p < 0.01$. Significance tests were adjusted using Bonferroni correction for three comparisons: No vs. Low, No vs. High, and Low vs. High.

Table 2

Correlation Between Outcome Measures in Study 1: Overall and by Levels of Uncertainty

	Overall (n = 495)				No Uncertainty (n = 165)				Low Uncertainty (n = 165)				High Uncertainty (n = 165)			
Measures	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1 Perceived Uncertainty	1				1				1				1			
2 Perceived Understandability	-0.15**	1			-0.22**	1			-0.21**	1			0	1		
3 Perceived Trust in Info.	-0.35***	0.46***	1		-0.36***	0.45***	1		-0.44***	0.39***	1		-0.21**	0.53***	1	
4 Perceived Effectiveness	-0.34***	0.32***	0.46***	1	-0.35***	0.41***	0.51***	1	-0.46***	0.26***	0.40***	1	-0.18*	0.26***	0.43***	1

Note: Correlations were collapsed across the three effect size metrics (i.e., months of progress, percentile gain, and raw score), as there was no evidence that the effect size metrics interacted with the level of uncertainty for any of the outcomes.

* p < 0.05, ** p < 0.01, *** p < 0.001

Robustness Analyses

Although assignment to conditions was randomized, we also examined whether the level of uncertainty influenced ratings on our four dependent variables after controlling for age, gender, teaching experience, familiarity with the statistical concept of effect size, familiarity with the KS2 Math test (i.e., the outcome measure in the vignettes), familiarity with the months of progress metric, self-reported mathematical ability, self-reported risk aversion, and online recruitment platform (Facebook or Prolific). The inclusion of these covariates did not affect the direction or significance of any of the effects found in our preregistered analysis (see section S1 for details). To examine the potential effect of attrition, we conducted two additional sensitivity analyses: (a) one including participants who completed some, but not all, of the survey, and (b) another using inverse probability weighting, assigning greater weight to participants less likely to have completed the survey (Gomila & Clark, 2020). Again, neither analysis changed the direction or significance of any of the effects found in our preregistered analysis (see Section S1 for details).

Discussion

The observed changes in perceived uncertainty and effectiveness, alongside the absence of an effect on perceived understandability, suggest that teachers do take account of uncertainty when it is reported, that research communicators should report this information, and that using a numerical range could be a viable option to do so. Conversely, the observed decrease in the perceived trustworthiness of the information when uncertainty is reported has more ambiguous implications. Different reasons could account for this effect, each with distinct implications for research communicators. One possibility is that the participants, not expecting any uncertainty, may have viewed uncertain estimates as less reliable and therefore less trustworthy. Another (not mutually exclusive) possibility is that participants may have interpreted uncertainty as a sign that the individuals providing the information lack competence, which in turn led to a reduction in trust. Past studies on risk communication have shown that a non-negligible proportion of participants respond in a manner consistent with this reasoning when presented with uncertainty (Johnson & Slovic, 1995; Johnson, 2003). These two possibilities both result in reduced trust, but, importantly, would have different implications for how research communicators should communicate. If the reduction of trust stems from the perceived unreliability of the estimate, communicators should aim to explain that uncertainty is an integral part of the scientific process and should be expected. If the issue relates to perceived incompetence, then communicators could consider incorporating signals of trustworthiness and competence in their reporting (see Jamieson et al., 2019 for ways to accomplish this).

Study 2

To better understand the impact of reporting uncertainty on perceived trustworthiness, we conducted a second study that explored how reporting uncertainty impacts (a) trust in the presented impact estimate and (b) trust in the source that wrote and published the description. Previous research indicates that presenting uncertainty may differentially affect these two types of trust (van der Bles et al., 2020).

We streamlined our instrument to maximize our chance of detecting any potential effects. Specifically, we compared only two conditions: no uncertainty mentioned and high uncertainty (omitting the low uncertainty condition from Study 1). Given the lack of interaction with the effect size metrics in Study 1, we only reported intervention effects in terms of months of progress (excluding percentile gain and raw test score). We added an attention question (per our preregistration, only participants who answered the question correctly were included in the analysis). Finally, we used this opportunity to replicate the effect of uncertainty on perceived effectiveness found in Study 1, as we felt that this effect is particularly important for research communicators to consider.

Methods

Participants

Five hundred teachers (92% female; Mean age: 38.58 years; SD: 8.76) were recruited using targeted advertising on Facebook to take part in an online survey. They had, on average, 12.24 years of teaching experience (SD: 8.21), mainly worked in the South (45%), Middle (28%), or North (27%) of England, and were mainly primary school (58%) or secondary school (24%) teachers. See Table S4 for a detailed description of the sample.

Unfortunately, due to an oversight, the survey did not record data from participants who did not complete it, preventing us from assessing the level of attrition. As per our pre-registration plan, only participants who completed the survey were included in the analysis. Importantly, as in Study 1, participants were similar across experimental conditions (see Table S4), and similar to census data (see Table S6). We also suspect attrition rate may have been lower than in Study 1, as the survey was shorter, potentially resulting in fewer incomplete responses.

Study Design

The study followed a between-subjects design in which we compared a group presented with the impact of an educational intervention with no uncertainty mentioned to a group receiving the same information but with a high level of uncertainty. We measured the impact of this manipulation on three dependent variables: (1) the perceived effectiveness of the intervention, (2) the perceived trustworthiness of the estimate of impact, and (3) the perceived trustworthiness of the people who wrote and published the description of impact.

Procedure

Participants were randomly assigned to one of the two uncertainty conditions. Instruction, vignettes, and presentation were identical to those used in Study 1, with the aforementioned difference that there was only one effect size metric (months of progress), two levels of uncertainty (not reported, and high) and three outcome measures: (1) perceived effectiveness of the intervention, (2) perceived trustworthiness of the estimate of impact, and (3) perceived trustworthiness of the source who wrote and published the description of impact. The questionnaire included the same demographic questions present in Study 1. Again, our analysis plan was preregistered prior to data collection. The analysis plan, along with the survey questionnaire, justification for sample size, analysis code, and data, can be found at: https://osf.io/35nwp/.

Results and Discussion

Figure 2 shows the mean ratings for the three dependent variables by level of uncertainty reported, and Table 2 shows their correlations.

Preregistered Analyses. We preregistered three analyses (i.e., three independent samples t-tests), one for each outcome measure.

Perceived effectiveness. We replicated the effect obtained in Study 1: presenting a range of uncertainty decreased perceived effectiveness $t(497.98) = 2.84$, $p = .005$, $d = 0.25$. Specifically, the intervention was perceived as being less effective when presented with uncertainty ($M = 4.70$, $SD = 1.18$), than with no uncertainty ($M = 5.00$, $SD = 1.18$).

Perceived trustworthiness in the estimate. Participants perceived the estimate as significantly less trustworthy when presented with uncertainty ($M = 3.74$, $SD = 1.25$) than without uncertainty ($M = 3.97$, $SD = 1.25$), $t(498) = 2.04$, $p = .041$, $d = 0.18$.

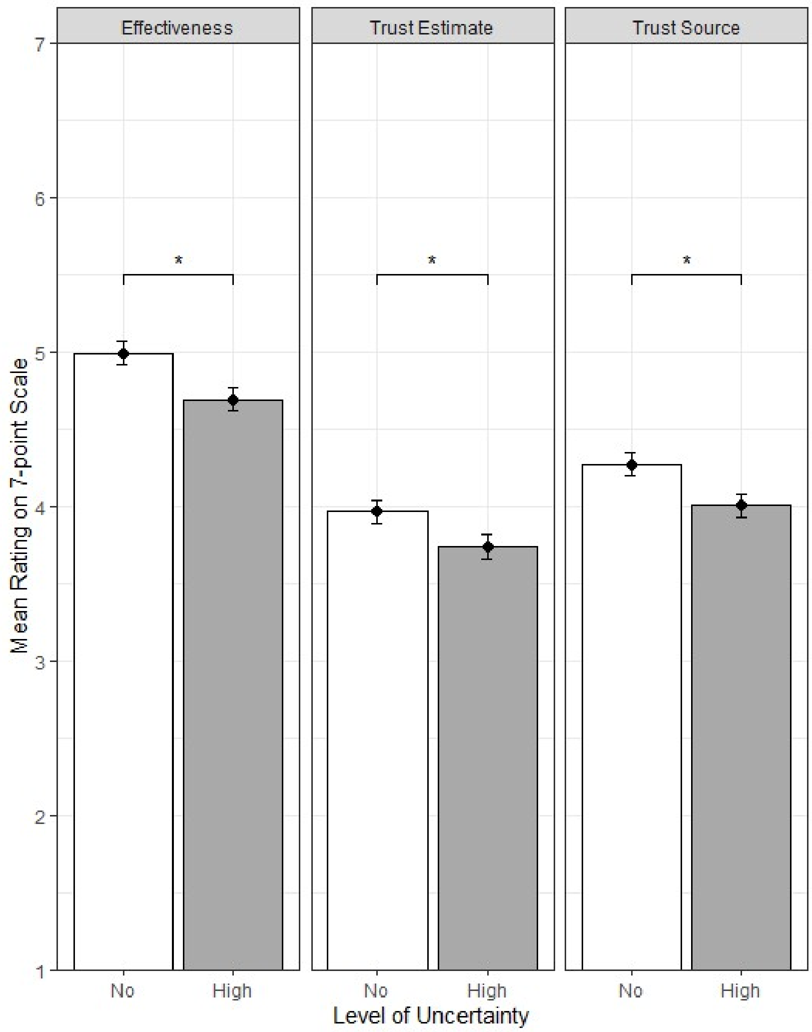
Perceived trustworthiness in the source. We also observed a similar decrease in trustworthiness in the people who wrote and published the description of impact. This was rated lower when the impact was presented with uncertainty ($M = 4.01$, $SD = 1.20$) than without uncertainty ($M = 4.28$, $SD = 1.14$), $t(496.75) = 2.57$, $p = .011$, $d = 0.23$. In addition to the effect of our manipulation, perceived trustworthiness of the source was also strongly related to perceived trust in the estimate (see Table 3). This is in line with the well-documented tendency for individuals who trust a source to also trust the information provided by that source (see Pornpitakpan, 2004, for a review).

Robustness Analyses

As in Study 1, we examined whether the effects observed persist after controlling for a range of covariates. Again, the inclusion of these covariates did not change the direction and significance of the effects found in our preregistered analysis (see section S2 for details).

Figure 2

Mean Ratings to the Three Dependent Variables by Level of Uncertainty Reported



Note: Error bars represent standard errors. Significance levels are indicated as follows: *$p < 0.05$, **$p < 0.01$.

Table 3

Correlation Between Outcome Measures in Study 2: Overall and by Levels of Uncertainty

	Overall (n = 500)			No Uncertainty (n = 250)			High Uncertainty (n = 250)		
Measures	1	2	3	1	2	3	1	2	3
1 Perceived Effectiveness	1			1			1		
2 Perceived Trust in estimate	0.50***	1		0.56***	1		0.42***	1	
3 Perceived Trust in source	0.36***	0.66***	1	0.37***	0.70***	1	0.34***	0.61***	1

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

General Discussion

There is a growing effort to inform teachers about the effects of various educational practices. However, the uncertainty associated with these effects is often very high, meaning that the effect of these practices will often deviate substantially from the impact estimates reported. While this suggests a need for research communicators to clearly report information about uncertainty, evidence from fields such as health sciences suggests that reporting uncertainty can be perceived negatively — a situation that could diminish teachers' engagement with research. We conducted two pre-registered studies to examine the impact of reporting uncertainty on five aspects of teachers' perceptions.

Perceived uncertainty.

We found that reporting uncertainty increased teachers' perceived uncertainty. This observation suggests teachers were responsive to our manipulation, but also suggests that they attribute more definitiveness to research findings than warranted. When no uncertainty was reported, teachers tended to expect less uncertainty than in the low-uncertainty conditions—a condition selected to represent a level of uncertainty considerably lower than what is typically observed in the field[8]. Given that uncertainty is frequently unreported or inadequately disclosed, this finding implies that teachers often underestimate uncertainty. This situation could lead teachers to misjudge the effect an intervention may have on their students, or to mistakenly believe that that two interventions with different point estimates must produce different impacts on their students. Ultimately, these oversights could erode teachers' views of the value of educational research.

Perceived understandability.

Presenting uncertainty involves providing additional information, numerical information in our case, a manipulation that could potentially reduce the perceived understandability of the information. Despite our sample size allowing us to detect relatively small effects, we found no evidence that reporting uncertainty had an impact on perceived understandability of the information presented. Considering that understandability is an important predictor of engagement (Davis, 1989), this is noteworthy. The lack of variation in perceived understandability also suggests that the other effects observed (on perceived uncertainty, effectiveness and trustworthiness) are unlikely to result from difficulty in processing the additional uncertainty information presented.

Perceived effectiveness of the intervention.

Although presenting a range of uncertainty emphasizes that the intervention effect could be lower or higher than the point estimate provided, our manipulation tended to *reduce* teachers' perceptions of an intervention's effectiveness. This response may reflect a conservative decision-making style, a tendency that would be understandable given the pressures teachers often face to

---

[8] As noted earlier, only 18% of the effects from rigorous, large-scale RCTs commissioned by the EEF have a 95% CI narrower than the range reported in the low uncertainty condition.

improve their students' achievement and the penalties associated with failing to do so. Similar decision-making patterns are documented in other occupations (such as physicians, e.g., Studdert et al., 2005). Our findings also mirror the upward bias documented in risk communication (e.g., Johnson & Slovic, 1998), where presenting uncertainty information often led participants to adopt a more cautious response—perceiving the reported risk to be higher in that context.

Considering the interest teachers have in the impact of educational interventions when deciding whether to implement a given intervention (Holland et al., 2019), our finding suggests that information about uncertainty has the potential to influence teachers' choices. Not reporting uncertainty denies teachers information they would likely use in their decision-making process, such as giving greater weight to more certain effects. Given the importance of this finding for decision-making, we aimed to replicate it in Study 2 and successfully did so.

Perceived trustworthiness.

While the impact on perceived uncertainty, perceived effectiveness, and the lack of impact on perceived understandability all make a strong case for the importance of reporting uncertainty in research communication, the negative impact observed on trustworthiness is likely to be a concern for research communicators. Trust is an important predictor of engagement (e.g., Kelton et al., 2008). Study 1 found that reporting uncertainty negatively impacts trust in the information presented, and Study 2 further investigating this effect showed a reduction in both (a) trust in the estimate presented, and (b) trust in the source that wrote and published the information.

The impact of reporting uncertainty on trust in the estimate may simply reflect the greater perceived uncertainty reported earlier — findings with more uncertainty (i.e., less reliability) are deemed less trustworthy. However, the reduction in trust in the source who wrote and published the information is concerning. Given that uncertainty is ubiquitous, one might think that a source reporting it would be seen as transparent and trustworthy. Our findings suggest otherwise. Perhaps not aware of the ubiquity and magnitude of uncertainty typical in the field, teachers may have attributed the presence of uncertainty to the incompetence of the communicators or to dishonest intentions. This finding is not unprecedented—Johnson and Slovic (1995), for example, found that a considerable number of participants (one-third of their sample) perceived an agency reporting risk with uncertainty as less competent. Moreover, this is not as incoherent as one may suspect: in many contexts, uncertainty often does reflect a lack of competence. For example, in a school setting, a student who claims that the product of 3 times 4 could be as high as 15 or as low as 9 is likely less competent than a student who confidently states that the answer is 12.

While our findings may not be entirely unexpected, they do somewhat contradict some recent studies, which show that presenting uncertainty, especially numerical uncertainty as in the present studies, has minimal impact (if any) on participants' trust in the numbers presented, and in the sources communicating those numbers (van der Bles et al., 2020). This discrepancy may be due to the different populations studied, namely teachers versus the general public, and the different contexts —

the impact of educational interventions versus government statistics. This variation underscores the need for teacher-specific research and suggests caution when generalizing existing findings to the teacher population.

Limitations

Readers may wonder if our effects are due to idiosyncratic features of our material. While possible, studies examining the impact of different presentation formats of numerical uncertainty (e.g., using a ± symbol rather than reporting the lower and upper bound) or the specific questions used to measure participants' perceptions (e.g., using the word “reliability” instead of “trustworthiness”) found no evidence that these variations influenced perceptions (e.g., van der Bles et al., 2020) — suggesting that any such effects, if they exist, are likely minimal. While our studies focus on numerical uncertainty, due to its relative clarity, one may also wonder if our findings generalize to more qualitative presentations of uncertainty, such as tier systems. Although this is an empirical question, research indicates that the negative effect on trustworthiness tends to be more pronounced when uncertainty is conveyed less transparently (such as verbally; van der Bles et al., 2020), suggesting these negative effects could be more severe. Readers may also wonder if the effects observed only arise when uncertainty reported is unusually high. While most effects in Study 1 only reached significance in the “high uncertainty” condition, this condition is by no means unrealistic. At the time of writing this, nearly 40% (i.e., 39.2%) of effects from rigorous large scale RCTs commissioned by the EEF have 95% confidence intervals wider than the range reported in our high uncertainty condition. We expect that this prevalence would also be greater in smaller, more typical, educational studies. It is also possible that our findings would have differed if the lower bound of the uncertainty range presented had been higher than 0. While this is possible, readers should keep in mind that such instances are relatively rare. For example, the majority of effects (> 77%) from large-scale RCTs commissioned by the EEF and NCEE had the lower boundary of their 95% CI at or below zero (Lortie-Forgues & Inglis, 2019). Finally, readers should note that we observed some level of attrition in Study 1 (16% post-randomization), and attrition could not be evaluated in Study 2. Nevertheless, robustness analyses for Study 1 suggest that the impact of attrition, if any, would be minimal. We also found no evidence that participants’ background characteristics differed between conditions in either study.

Implications

Our findings suggest that teachers would use uncertainty information in their decision-making, but that reporting this information would result in reduced trustworthiness (both in the estimate and in the people who communicate the information) — effects that could lessen teachers’ engagement with research. This presents a dilemma: research communicators aiming to assist teachers may inadvertently undermine their own objectives. We believe that research communicators should still report uncertainty, as failing to do so may lead teachers to misestimate the effect of educational interventions, which could, over time, erode their trust in research. In this context, a viable compromise could be to communicate uncertainty while concurrently taking steps to mitigate its impact on trust. This could include strategies like disclosing competing interests or indicating that an independent review has been conducted (for additional methods of signaling trust, see Jamieson et al., 2019).

Our findings also have implications for research evaluators. The fact that most effects in Study 1 were only found in the highest uncertainty condition suggests that teachers may be sensitive to the magnitude of the uncertainty presented. This indicates that research evaluators concerned about their intervention being perceived as less effective should aim to estimate their effects with greater precision by, for example, recruiting larger samples.

Initial teacher training could also benefit from emphasizing uncertainty as a normal part of education and defining what constitutes reasonable uncertainty. This may not be as difficult as one may think. Our findings suggest that teachers are aware that educational effects are, to some extent, uncertain — indeed, the average uncertainty rating when no uncertainty was reported was not at the lowest point of the rating scale (see Figure 1). Moreover, uncertainty, at least the way we presented it, was not perceived as difficult to understand. It is possible, therefore, that helping teachers recognize that the uncertainty of educational effects is larger than what they expect — perhaps through a short, one-time debiasing training such as those described by Morewedge et al. (2015) or Sellier et al. (2019) — could be effective. Familiarizing teachers with the typical uncertainty in education research may lead them to view its reporting as a sign of trustworthiness, possibly reversing our observed effects. Teachers may even come to a point where they request that the uncertainty of an effect be clearly communicated to them.

Finally, our findings have implications for researchers interested in improving research communication. Knowing that teachers consider reported uncertainty opens up a range of important research questions. How do teachers perceive less transparent ways of reporting uncertainty, such as tier systems? Could teachers benefit from being presented with two representations of uncertainty (e.g., both numerical and qualitative), as is sometimes recommended (e.g., GRADE working group [Guyatt et al., 2008])? How do teachers interpret reported uncertainty? Considering the ubiquity and magnitude of uncertainty in education research and finding ways to ensure that teachers engage with this information in their decision-making should be seen as a priority.

References

American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). Washington, DC: American Psychological Association.

Ayaita, A., & Stürmer, K. (2020). Risk aversion and the teaching profession: An analysis including different forms of risk aversion, different control groups, selection and socialization effects. *Education Economics*, *28*(1), 4-25.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. John Wiley & Sons.

Budescu, D. V., Weinberg, S., & Wallsten, T. S. (1988). Decisions based on numerically and verbally expressed uncertainties. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(2), 281.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319-340.

Department for Education. (2019). National curriculum assessments: Key stage 2, 2019 (revised). GOV.UK. https://www.gov.uk/government/statistics/national-curriculum-assessments-key-stage-2-2019-revised

Education Endowment Foundation. (2023). Teaching and learning / Early years toolkit guide. Toolkit Guide. https://d2tic4wvo1iusb.cloudfront.net/production/documents/toolkit/Toolkit_guide_v1.2_-_2023.pdf?v=1712676551

Education Endowment Foundation. (2019). Classification of the security of findings from EEF evaluations. https://d2tic4wvo1iusb.cloudfront.net/documents/evaluation/peer-review-process/Classifying_the_security_of_EEF_findings_2019.pdf

Erev, I., & Cohen, B. L. (1990). Verbal versus numerical probabilities: Efficiency, biases, and the preference paradox. *Organizational behavior and human decision processes*, *45*(1), 1-18.Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Routledge.

Gomila, R., & Clark, C. S. (2022). Missing data in experiments: Challenges and solutions. *Psychological Methods, 27*(2), 143.

Guyatt, G. H., Oxman, A. D., Kunz, R., Vist, G. E., Falck-Ytter, Y., & Schünemann, H. J. (2008). What is “quality of evidence” and why is it important to clinicians?. *Bmj*, 336(7651), 995-998.

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for research in mathematics education*, *21*(1), 33-46.

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2023). *Cochrane Handbook for Systematic Reviews of Interventions* (Version 6.4) [Updated August 2023]. Cochrane. https://www.training.cochrane.org/handbook

Higgins, S., Katsipataki, M., Villanueva Aguilera, A. B., Dobson, E., Gascoine, L., Rajab, T., ... & Uwimpuhwe, G. (2022). The Teaching and Learning Toolkit: Communicating research evidence to inform decision-making for policy and practice in education. *Review of Education*, *10*(1), e3327.

Hollands, F., Pan, Y., & Escueta, M. (2019). What is the potential for applying cost-utility analysis to facilitate evidence-based decision making in schools?. *Educational Researcher*, *48*(5), 287-295.

Hultcrantz, M., Rind, D., Akl, E. A., Treweek, S., Mustafa, R. A., Iorio, A., ... & Guyatt, G. (2017). The GRADE Working Group clarifies the construct of certainty of evidence. *Journal of clinical epidemiology, 87*, 4-13.

Jamieson, K. H., McNutt, M., Kiermer, V., & Sever, R. (2019). Signaling the trustworthiness of science. *Proceedings of the National Academy of Sciences*, *116*(39), 19231-19236.

Johnson, B. B., & Slovic, P. (1995). Presenting uncertainty in health risk assessment: Initial studies of its effects on risk perception and trust. *Risk analysis*, *15*(4), 485-494.

Johnson, B. B., & Slovic, P. (1998). Lay views on uncertainty in environmental health risk assessment. *Journal of Risk Research*, *1*(4), 261-279.

Johnson, B. B. (2003). Further notes on public response to uncertainty in risks and science. Risk Analysis: An International Journal, 23(4), 781-789.

Kassambara A. (2023). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests* (R package version 0.7.2) [computer software]. Retrieved from https://cran.r-project.org/web/packages/rstatix/index.html

Kelton, K., Fleischmann, K. R., & Wallace, W. A. (2008). Trust in digital information. Journal of the American Society for Information Science and Technology, 59(3), 363-374.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational researcher*, *49*(4), 241-253.Lipkus, I. M., Klein, W. M., & Rimer, B. K. (2001). Communicating breast cancer risks to women using different formats. *Cancer Epidemiology Biomarkers & Prevention*, *10*(8), 895-898.

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., ... & Busick, M. D. (2012). Translating the Statistical Representation of the Effects of Education Interventions into More Readily Interpretable Forms. *National Center for Special Education Research*.

Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned?. *Educational Researcher*, *48*(3), 158-166.

Lortie-Forgues, H., Sio, U. N., & Inglis, M. (2021). How should educational effects be communicated to teachers?. *Educational Researcher*, *50*(6), 345-354.

Peters, E., Hibbard, J., Slovic, P., & Dieckmann, N. (2007). Numeracy skill and the communication, comprehension, and use of risk-benefit information. *Health affairs*, *26*(3), 741-748.

Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of applied social psychology, 34*(2), 243-281.

Schapira, M. M., Nattinger, A. B., & McHorney, C. A. (2001). Frequency or probability? A qualitative study of risk communication formats used in health care. *Medical Decision Making*, *21*(6), 459-467.

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *Journal of Pharmacology and pharmacotherapeutics, 1*(2), 100-107.

Sellier, A. L., Scopelliti, I., & Morewedge, C. K. (2019). Debiasing training improves decision making in the field. *Psychological science*, *30*(9), 1371-1379.

Simpson, A. (2017). The misdirection of public policy: Comparing and combining standardised effect sizes. *Journal of Education Policy*, *32*(4), 450-466.

Studdert, D. M., Mello, M. M., Sage, W. M., DesRoches, C. M., Peugh, J., Zapert, K., & Brennan, T. A. (2005). Defensive medicine among high-risk specialist physicians in a volatile malpractice environment. *Jama*, *293*(21), 2609-2617.

van der Bles, A. M., van der Linden, S., Freeman, A. L., & Spiegelhalter, D. J. (2020). The effects of communicating uncertainty on public trust in facts and numbers. *Proceedings of the National Academy of Sciences*, *117*(14), 7672-7683.

van der Bles, A. M., Van Der Linden, S., Freeman, A. L., Mitchell, J., Galvao, A. B., Zaval, L., & Spiegelhalter, D. J. (2019). Communicating uncertainty about facts, numbers and science. *Royal Society open science*, *6*(5), 181870.

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). https://ies.ed.gov/ncee/wwc/Handbooks.