

Teacher’s Reasons for Trust and Distrust in Scientific Evidence: Reflecting a “Smart But Evil” Stereotype?

Tom Rosman 

Leibniz Institute for Psychology Information (ZPID), Germany

Samuel Merk 

PH Karlsruhe, Germany

We investigate in-service teachers’ reasons for trust and distrust in educational research compared to research in general. Building on previous research on a so-called “smart but evil” stereotype regarding educational researchers, three sets of confirmatory hypotheses were preregistered. First, we expected that teachers would emphasize expertise—as compared with benevolence and integrity—as a stronger reason for trust in educational researchers. Moreover, we expected that this pattern would not only apply to educational researchers, but that it would generalize to researchers in general. Furthermore, we hypothesized that the pattern could also be found in the general population. Following a pilot study aiming to establish the validity of our measures (German general population sample; $N = 504$), hypotheses were tested in an online study with $N = 414$ randomly sampled German in-service teachers. Using the Bayesian informative hypothesis evaluation framework, we found empirical support for five of our six preregistered hypotheses.

Keywords: *epistemic trust, trust in researchers, teacher education, evidence-based teaching*

THE integration of research results into teachers’ classroom practice is becoming ever more important (e.g., Southerland et al., 2016). However, several studies show that both preservice and in-service teachers tend to prefer conventional wisdom and practically derived knowledge sources over scientific evidence (Bråten & Ferguson, 2015; Cramer, 2013; Parr & Timperley, 2008). This has potentially far-reaching consequences: Teachers not knowing the science behind their profession may succumb to misconceptions on teaching and learning, which, in turn, may lead to dysfunctional educational practices. For example, they may waste time developing teaching materials tailored to individual students’ learning styles (the “learning style” myth has long been debunked; Kirschner, 2017; Pashler et al., 2008), or they may recommend grade retention for lower performing students (there is no evidence of academic benefit for retained students; Hughes et al., 2018; Jimmerson, 2001).

Several authors have argued that this underreliance on research knowledge might be grounded in beliefs about the sources of educational knowledge (e.g., Bråten & Ferguson, 2015; Merk et al., 2017; Sjølie, 2014). This is because directly evaluating the validity and robustness of educational research is hard for teachers, as it often requires a considerable amount

of background knowledge in research methodology (Hendriks et al., 2021). Therefore, to “acquire reliable and useable information about learning and teaching, (pre-service) teachers must instead be able to identify and evaluate knowledgeable and trustworthy sources of information—that is, to figure out whom to believe” (Hendriks et al., 2021, p. 2). Keeping in mind the importance of such “second-hand evaluations” (i.e., figuring out “whom to believe” instead of “what is true”; Bromme et al., 2015), we know from social psychology that members of one’s in-group are seen as more positive compared with individuals from outside one’s group (in-group bias; Mullen et al., 1992). Therefore, teachers might rate fellow teachers as more trustworthy compared with researchers. Furthermore, considering the findings that teachers view researchers as competent but self-interested (Critchley, 2008), and that researchers are generally ascribed less warmth (i.e., benevolence) but similar competence (i.e., expertise) compared with teachers (Fiske & Dupree, 2014), one might suppose that this is especially true for researchers’ perceived benevolence and integrity.

In line with such arguments, Merk and Rosman (2019) found what they subsequently called the “smart but evil” stereotype: In two experimental studies, they showed that pre-service teachers view educational researchers’ expertise as



significantly higher than their integrity and benevolence compared with their views on practitioners (i.e., in-service teachers). While it is still disputed whether comparably lower—but still high—integrity and benevolence ascriptions may indeed be labeled as “evil” (Hendriks et al., 2021), Hendriks et al. (2021) found additional (partial) support for a smart but evil pattern in preservice teachers: When following the epistemic aim of acquiring theoretical explanations in a teaching context, they saw researchers as having more expertise and integrity but less benevolence compared with practitioners. In contrast, when their participants were looking to gain practical advice for everyday school life, teachers were seen, compared with researchers, as having altogether higher expertise, integrity, and benevolence. In other words, researchers were generally less trusted than teaching practitioners, with the exception of their expertise and integrity in providing theoretical explanations.

Taken together, these studies provide evidence for the existence of a smart but evil pattern. As impaired epistemic trust might lead to teachers disregarding empirical evidence in their teaching, close attention to such findings is warranted. To date, however, the existing evidence does not allow us to specify the extent to which the pattern is specific for the domain of educational research (Merk & Rosman, 2019), especially since the aforementioned study by Fiske and Dupree (2014) found a similar pattern in a general population sample. Moreover, current studies on epistemic trust (including the works by Merk & Rosman, 2019, and Hendriks et al., 2021) do not allow the delineation between trust and distrust since they simply operationalize distrust as being the opposite of trust (which is not necessarily true; e.g., Lewicki et al., 1998). Furthermore, it is not yet clear whether the pattern only applies to *comparisons* between researchers and practitioners, or whether the smart but evil pattern also manifests itself in the mean differences between researchers’ expertise, integrity, and benevolence. Finally, considering the role of teachers as multipliers of knowledge and beliefs, the studies’ focus on preservice teachers narrows the generalizability of the findings. Hence, in the present article, we strive to replicate the smart but evil pattern and address the mentioned shortcomings by (1) analyzing reasons for trust and distrust in domains of differing granularity (i.e., educational research *and* research in general), (2) investigating trust and distrust as separate concepts, (3) employing a different methodological approach, and (4) using a larger and more heterogeneous sample (i.e., in-service teachers throughout Germany). The findings outlined here provide a foundation for future research focusing on the differentiated analysis of trust and distrust in (educational) researchers.

The Concept of Epistemic Trust

Epistemic trust is defined as the amount of trust that individuals ascribe to a specific knowledge source such as, for

example, educational researchers. As Hendriks et al. (2016) point out, this amount of epistemic trust depends on the epistemic trustworthiness of a source—in other words, on specific information features of the source that make it more or less trustworthy. For example, teachers will likely ascribe more trust to a renowned professor compared with an undergraduate student in his or her first semester. From a psychological perspective, this is because individuals use specific source information features to gauge a source’s epistemic trustworthiness (Hendriks et al., 2016; Landrum et al., 2015). Several researchers have proposed to distinguish between three dimensions that individuals use when evaluating this trustworthiness using source information features: expertise, integrity, and benevolence (e.g., Hendriks et al., 2015, 2016; Mayer et al., 1995). A source with high expertise (or ability; Mayer et al., 1995) is highly skilled and qualified within a particular domain (Hendriks et al., 2015, 2016). A source with high integrity is honest and adheres to recognized standards (e.g., transparency and openness) in his or her field (Hendriks et al., 2015). Finally, a benevolent source has good intentions and acts for the greater good of others—in contrast to someone who is only interested in his or her personal benefit (Hendriks et al., 2016; Mayer et al., 1995). Here, we must mention the similarities that expertise and benevolence share with the stereotype content dimensions of competence (e.g., competent, capable, intelligent) and warmth (e.g., sincere, friendly, well-intentioned) suggested by Fiske et al. (2002).

Established measurement instruments on epistemic trustworthiness (e.g., the Münster Epistemic Trustworthiness Inventory [METI]; Hendriks et al., 2015) often operationalize trust and distrust, along the dimensions outlined above, as opposite ends of three continuous variables (i.e., expertise, benevolence, and integrity; see also Saunders & Thornhill, 2004). However, a growing body of theories and empirical studies suggest that (1) trust and distrust may coexist or (2) that they may be separate constructs (Bigley & Pearce, 1998; Ou & Sia, 2010; Saunders et al., 2014; Saunders & Thornhill, 2004; Sitkin & Roth, 1993). With regard to the latter, *distrust* arises when “fundamental values are violated, and perceived trustworthiness is undermined across contexts” (Sitkin & Roth, 1993, p. 370). In contrast, violations in *trust* are seen as specific to a particular context, which is why distrust, because of its higher generality, would be more persistent and harder to change than reduced trust (Sitkin & Roth, 1993). Regarding the possible coexistence of trust and distrust, findings from social psychology suggest that humans may well express attitudes of positive valence and negative valence simultaneously (Cacioppo et al., 1997). More specifically, Cacioppo et al. (1997) argue that “a stimulus may vary in terms of the strength of positive evaluative activation and the strength of negative evaluative activation it evokes” (p. 3). Therefore, it is possible that certain stimuli evoke a strong activation of both positive *and* negative evaluative

processes, thus resulting in attitude ambivalence (Cacioppo et al., 1997). While we are not aware of any corresponding studies, such arguments are easily transferred to trust in (educational) researchers. For example, teachers may very well trust the Programme for International Student Assessment team's expertise in large-scale data analysis, while at the same time exhibiting a more abstract and general distrust in their ability to derive adequate conclusions on students' competencies.

In line with such arguments, it is not surprising that the 2017 and 2018 Science Barometer ("Wissenschaftsbarometer"), a representative survey of German citizens' attitudes toward science and research (Wissenschaft im Dialog/Kantar Emnid, 2017, 2018), not only included items assessing participants' individual trust in researchers but also items assessing why they *distrusted* researchers.

The Smart but Evil Pattern

In 2019, Merk and Rosman investigated whether pre-service teachers differ in the amount of epistemic trustworthiness they ascribe to different sources of educational knowledge. Specifically, they confronted their participants with short texts from the educational domain (e.g., on the prevalence of bullying in schools) that were experimentally manipulated with regard to their alleged source (practitioner vs. scientific study) while remaining invariant in content. For example, in one text version, a practitioner (i.e., a teacher) reported on his or her experiences regarding the prevalence of bullying, whereas another text version contained the same information, but was framed as a report of an empirical study written by researchers (Merk & Rosman, 2019). After reading each text, participants were asked to rate the epistemic trustworthiness of the texts' authors on the three METI dimensions (expertise, integrity, and benevolence; Hendriks et al., 2015). In two experimental studies, Merk and Rosman (2019) found what they later coined as the smart but evil stereotype: Educational researchers were seen as having less integrity and benevolence but, at the same time, as having more expertise compared with practitioners. Furthermore, as outlined above, Hendriks et al. (2021) found that pre-service teachers with the epistemic aim of receiving theoretical explanations regarded researchers as having more expertise and integrity but less benevolence as compared with practitioners, thus lending further support for a smart but evil pattern (but note their diverging findings on the integrity dimension). Nevertheless, one should keep in mind that such findings of researchers being perceived as more smart but evil *compared with practitioners* (e.g., Hendriks et al., 2021; Merk & Rosman, 2019) do not imply that educational researchers are seen as smart but evil *per se*, especially when considering the rather high means that were found for the METI dimensions across all experimental conditions.

The Generalizability of Epistemic Trust Across Different Scientific Domains

In their theoretical framework on learning to trust and trusting to learn, Landrum et al. (2015) suggest that learners will generalize their beliefs about whom to trust by referring to "an unfamiliar individual's domain of expertise to make inferences about what he or she is likely to know" (p. 110). As both educational researchers and researchers in general stem from the "science" domain, a certain amount of generalizability of epistemic trust from educational research to other scientific domains (and vice versa) is thus likely. Further support for this assumption comes from the field of epistemic beliefs. In fact, as evidenced by a recent study (Merk & Rosman, 2019), epistemic beliefs and epistemic trust share a certain conceptual overlap—while epistemic beliefs focus on how people think about *knowledge* itself (Hofer & Pintrich, 1997), epistemic trust denotes how individuals evaluate the expertise, benevolence, and integrity of a specific *knowledge source* (Hendriks et al., 2016). Therefore, the Theory of Integrated Domains in Personal Epistemology (TIDE; Merk et al., 2018; Muis et al., 2006) provides a foundation for the generalizability of epistemic trust across domains. In fact, the TIDE framework suggests that epistemic beliefs from one domain (e.g., biology-specific epistemic beliefs) influence both more general (e.g., academic) and more specific (e.g., topic-specific) epistemic beliefs, and several studies have found empirical evidence for such predictions (e.g., Merk et al., 2018; Muis et al., 2006).

The Present Study

The present study, which was preregistered at PsychArchives (Rosman & Merk, 2020), builds on Merk and Rosman's (2019) work and extends it with regard to the aspects outlined above. To do so, we use data from a cross-sectional survey asking German in-service teachers about three reasons (expertise, benevolence, and integrity) for their epistemic trust and distrust. Participating teachers were asked to provide two responses (on Likert-type scales)—one regarding researchers in general, and one regarding educational researchers.

In a first step, we investigate these reasons with regard to the latter group of educational researchers. Our underlying assumption is that if the smart but evil pattern is present in teachers, it will impact their explanations on why they trust or distrust educational researchers. More specifically, when justifying their trust in educational researchers, we expect that our participants will more strongly refer to expertise-related reasons rather than provide explanations focusing on educational researchers' benevolence and integrity. In contrast, to justify their distrust in educational researchers and in line with the smart but evil pattern, we expect our participants to more strongly refer to benevolence- and integrity-related reasons compared with expertise-related reasons. To

conceptually replicate the smart but evil pattern in in-service teachers, we therefore posit the following confirmatory hypotheses:

Hypothesis 1: Concerning their reasons for trusting educational researchers, teachers will score higher on expertise-related reasons compared with benevolence-related reasons and integrity-related reasons (H_1^1).

Hypothesis 2: Concerning their reasons for distrusting educational researchers, teachers will score lower on expertise-related reasons compared with benevolence-related reasons and integrity-related reasons (H_2).

Second, we will test whether these hypothesized effects are specific to educational research or whether they apply to research in general. Tentatively, one may expect the latter. In fact, a generalizability of the reasons for trust and distrust across domains would be in line with the aforementioned trust research and epistemic beliefs frameworks (e.g., Landrum et al., 2015; Merk et al., 2018; Muis et al., 2006). The empirical results on the reasons for trust and distrust in the Science Barometer (Wissenschaft im Dialog/Kantar Emnid, 2017, 2018) further support this assumption: As outlined above, the Science Barometer included three trust items and three distrust items pertaining to researchers in general. The response patterns obtained for these items—at least descriptively—speak in favor of a preference for expertise-related reasons over reasons of benevolence and integrity regarding trust in researchers in general (and for the opposite pattern regarding distrust; Könneker, 2018, 2020; Wissenschaft im Dialog/Kantar Emnid, 2017). To further investigate this generalizability aspect, we will analyze whether the reasons for trust and distrust suggested in Hypotheses 1 and 2 differ when teachers are requested to provide their responses with regard to researchers in general. We justify the importance of these analyses as follows: If teachers' smart but evil patterns generalize from educational researchers to researchers in general, this would suggest that teachers are not specifically biased toward educational researchers, thus making it easier to adopt interventions for rebuilding trust in science (e.g., increased transparency; Bachmann et al., 2015) to the educational context. In line with our reasoning above, we suggest the following confirmatory hypotheses:

Hypothesis 3: Concerning their reasons for trusting researchers in general, teachers will score higher on expertise-related reasons compared with benevolence-related reasons and integrity-related reasons (H_3).

Hypothesis 4: Concerning their reasons for distrusting researchers in general, teachers will score lower on expertise-related reasons compared with benevolence-related reasons and integrity-related reasons (H_4).

Testing the similarity of and differences between belief configurations across domains has a certain tradition in epistemic beliefs research because of its implications for theory building (e.g., Merk et al., 2018). More specifically, the TIDE framework (see above; Merk et al., 2018) predicts that beliefs about knowledge regarding one domain generalize to more general beliefs about scientific knowledge. Thus, examining the magnitude of differences between smart but evil patterns for educational researchers and researchers in general might lend further support to such assumptions. Furthermore, if we were to show that such differences are rather small, this would make it even easier to adopt more general interventions on rebuilding trust to the context of educational research. Hence, if Hypotheses 1 and 3 or 2 and 4 are both significant, we will conduct additional exploratory analyses on a within-person level to examine the magnitude of differences between reasons for (dis)trusting educational researchers and researchers in general.

As a third set of hypotheses, we will investigate, using additional data (2018 Science Barometer; Wissenschaft im Dialog/Kantar Emnid, 2018), whether the general populations' explanations for their trust and distrust in researchers also reflect the patterns outlined in the previous hypotheses. While this set of hypotheses does not directly relate to the teacher educational context, it allows us to test whether the patterns suggested in Hypotheses 3 and 4 are specific for teachers or whether they generalize to a comparable population of German nonteachers. As outlined above, the 2017 and 2018 Science Barometer data support the latter assumptions on a descriptive level (Wissenschaft im Dialog/Kantar Emnid, 2017, 2018). In the present study, we will investigate whether this still holds true when matching the 2018 Science Barometer sample to our teacher sample with regard to age and socioeconomic status (SES; e.g., education and interest in politics, science, and sports). Such analyses are important because they allow researchers to determine whether teachers are a special subgroup who have distinct beliefs about researchers or whether their patterns of epistemic trust are largely similar to those of the general population. If the latter were true, this would, again, allow to adopt more general interventions on rebuilding trust to teachers without much effort—which would certainly be good news. We thus posit the following hypotheses:

Hypothesis 5: Concerning their reasons for trusting researchers in general, a general population sample matched for age and SES will score higher on expertise-related reasons compared with benevolence-related reasons and integrity-related reasons (H_5).

Hypothesis 6: Concerning their reasons for distrusting researchers in general, a general population sample matched for age and SES will score lower on expertise-related reasons compared with benevolence-related reasons and integrity-related reasons (H_6).

While this set of hypotheses does not directly relate to the teacher education context, it allows us to test whether our predictions are specific for teachers or whether they generalize to a comparable population of German nonteachers. For this research question, estimating the magnitude of potential differences between teachers and the general population in their reasons for trust in researchers might be particularly interesting because if we find distinct trust patterns for teachers as compared with the general population, it is important to uncover the extent of these differences. If the analyses determine that both Hypotheses 3 and 5 are significant, we will conduct corresponding exploratory analyses using a data set that combines our data with the matched 2018 Science Barometer data.

Method

Pilot Study

To draw valid comparisons between a teacher sample and the general population sample from the 2018 Science Barometer (Wissenschaft im Dialog/Kantar Emnid, 2018), it is important that the trust and distrust measurements are identical across both studies. However, the 2018 Science Barometer draws on nonvalidated single items, thus making it impossible to assess their psychometric properties. Moreover, it is important to note that all our hypotheses focus on differences between expertise, benevolence, and integrity as *reasons* for trust and distrust—which complicates making inferences about the objective (or absolute) levels of trust and distrust. To address these issues, we additionally conducted a pilot study² aiming at testing the validity of our items.

For this study, we recruited a German general population sample of $N = 504$ adults (252 female; 11% 18–24 years; 15% 25–31 years; 15% 32–38 years; 13% 39–45 years; 15% 46–52 years; 18% 53–59 years; 14% 60–66 years) using an online sample provider. In addition to a number of survey questions unrelated to the present study, participants were asked about their reasons for trusting and distrusting researchers in general, using the exact same items that were included in the 2017 and 2018 Science Barometer (Wissenschaft im Dialog/Kantar Emnid, 2017, 2018). Participants responded to two items per dimension (expertise, benevolence, integrity) on scales with five response categories (ranging from “do not agree” to “fully agree”). Specifically, for a given dimension, one item addressed reasons to trust researchers and the second item addressed reasons to distrust researchers (see Table 1). All scales included a “don’t know” option which was treated as missing data and dealt with using casewise deletion.

To validate these items, we additionally administered the METI (Hendriks et al., 2015) to assess researchers’ expertise, integrity, and benevolence using a 14-item adjective-based semantic differential with five response categories (e.g., “incompetent-competent” as an indicator of expertise).

Furthermore, we included six reworded items of the “reasons for trust” indicators described above. These items as well as their introduction had been stripped of all content pertaining to *reasons* for trust, thus directly asking participants about their epistemic trust in researchers in general (e.g., “Because researchers are experts in their field” was replaced by “Researchers are experts in their field,” and the notion, “There are several reasons for trusting researchers” was removed from the introduction). The order of the survey pages including the original and reworded items was balanced out, and the METI was administered between the two variants.

To investigate³ these data, we first tested the factorial validity of the METI. As expected, the theoretically proposed three-dimensional confirmatory factor analysis model outperformed competing models and showed good fit indices referring to classical benchmarks like those proposed by Hu and Bentler (1999). Subsequently, we conducted a multitrait-multimethod analysis (Eid, 2000) to assess the construct validity of the Science Barometer items. Specifically, we fitted a correlated trait/uncorrelated methods model, inspected the fit indices, and tested if the standardized loadings of the Science Barometer items on the corresponding dimension were substantial ($> .30$) and largely equivalent (absolute difference of standardized loadings $< .15$) to the loadings of the reworded items using approximative adjusted fractional Bayes factors for structural equation modeling (Gu et al., 2019). The theoretically expected correlated traits/uncorrelated methods model showed good fit measures for the Science Barometer items measuring trust, $\chi^2(df) = 424.68(147)$, Tucker–Lewis index (TLI) = 0.951, comparative fit index (CFI) = 0.962, root-mean-square error of approximation (RMSEA) = 0.061, standardized root-mean-square residual (SRMR) = 0.029 as well as for those measuring distrust, $\chi^2(df) = 471.01(147)$, TLI = 0.938, CFI = 0.952, RMSEA = 0.066, SRMR = 0.040, which we view as strong evidence for the convergent validity of these items. This was corroborated by the strong evidence provided by the Bayes factors (i.e., 10 of the 12 standardized trait factor loadings of the single items were substantial with a $BF_u > 4$; the two small factor loadings were lack of expertise as reason for distrust and the corresponding rewording), and all Science Barometer items loaded equivalent to their corresponding rewordings ($BF_u > 4$).

Participants and Procedure

All study procedures (except the pilot study) were preregistered at PsychArchives (Rosman & Merk, 2020). Hypotheses were tested in a sample of German in-service school teachers from schools all over Germany. Participants were recruited by means of a professional opinion research service (forsa). All participants who agreed to participate in the study were then then directed to an online survey where the study data were collected. In total, $N = 414$ in-service teachers completed the data collection (67% female; age: $M = 47.7$; $SD = 10.8$; teaching experience: $M = 17.4$ years,

TABLE 1

Trust and Distrust Items for Each of the Three Dimensions of Epistemic Trust (According to Wissenschaft im Dialog/Kantar Emnid, 2017; Translation by the Authors)

	Reasons for trust	Reasons for distrust
Expertise	Because researchers are experts in their field.	Because researchers frequently make mistakes.
Integrity	Because researchers work according to rules and standards.	Because researchers often adapt results to their own expectations.
Benevolence	Because researchers do their research in the interest of the public.	Because researchers are strongly dependent on their funders.

Note. The terms “researchers” were replaced with “educational researchers” in the second part of the questionnaire. Introduction: “There are several reasons for [trusting/distrusting] researchers. Please indicate your agreement to each of the following reasons.”

$SD = 11.1$; 26.1% primary schools, 29.5% grammar schools; 79% former West Germany).

As the items analyzed for the present study were part of a larger survey with experimental elements, the sample size was already determined by the requirements of these elements. As we planned to test our hypotheses using approximated adjusted Bayes factors for informative hypotheses (Gu et al., 2018; Hoijtink, 2011), we ran a Bayes factor design analysis (Schönbrodt & Wagenmakers, 2018) for this fixed N to estimate the “probability of achieving a research goal” (i.e., statistical power; Kruschke, 2010, p. 658) of our analyses. We thereby specified the smallest effect size of interest to $d = .30$. As can be seen in the Bayes factor design analysis documentation in our preregistration (Rosman & Merk, 2020), the planned decision procedure leads, according to our simulations, very rarely to “inconclusive” or “wrong” (false positive or false negative) results (see preregistration for more details; Rosman & Merk, 2020).

Design and Materials

Even though later parts of the survey drew on an experimental design, all data used for the present study were collected before assignment to any experimental groups (Rosman & Merk, 2020). Each participant thus received an identical set of materials. All materials were administered in German language; the examples below have been translated.

At the beginning of the survey, some covariates were measured (mostly single items; i.e., interest in science, politics, and sports; self-reported scientific literacy; beliefs about science; general trust in science). These covariates were taken from the 2018 Science Barometer, and they were required for the investigation of other research questions (which were significantly different from those of the present article). Subsequently, participants were asked about their reasons for trusting and distrusting researchers in general using the Science Barometer questions also employed in our pilot study (see Table 1). After responding to other covariates (interest in educational research; self-reported scientific literacy regarding educational research; beliefs about educational research; general trust in educational research), participants were asked about

their reasons for trusting and distrusting educational researchers. These questions were again identical to the Science Barometer questions with the exception that they related to educational researchers rather than researchers in general. Specifically, we exchanged the term “researchers” with “educational researchers.” Furthermore, a brief definition of “educational research” was provided to reduce the risk of bias by participants conceptualizing educational researchers in different ways (“Educational research focuses on the theory and practice of education and pedagogy. Subdisciplines are, among others, pedagogy, educational psychology, educational economics, and educational sociology”). Again, participants responded on 5-point Likert-type scales with an additional “don’t know” option.

Data Analysis

Our preregistered confirmatory hypotheses were tested using Bayesian Informative Hypothesis Evaluation (i.e., the so-called bain framework; Hoijtink, 2011) by means of the R-package bain (Gu et al., 2019). Through the estimation of Bayes factors, this approach generates relative evidence on how much more likely the current data are to be observed under a specific hypothesis H_i compared with another

hypothesis H_j $\left(BF_{ij} = \frac{p(Data | H_i)}{p(Data | H_j)} \right)$. Furthermore, the

bain framework allows researchers to specify hypotheses that contain equality and inequality constraints as well as order constraints. For example, in an analysis of variance (ANOVA) context, one may be interested in comparing the means of a baseline group (μ_{base}), a group that received an intervention A (μ_A), a group that received an intervention B (μ_B), and a group that received a combined intervention AB (μ_{AB}). However, the classical frequentist procedure only gains evidence against the null hypothesis $H_0 : \mu_{base} = \mu_A = \mu_B = \mu_{AB}$. In contrast, the bain framework is able to provide relative evidence for or against much more general and therefore more informative (Schnell et al., 2008) hypotheses such as $H' : 0 = \mu_{base} < \mu_A = \mu_B < \mu_{AB}$ or $H'' : 0 = \mu_{base} < \mu_A < \mu_B < \mu_{AB}$.

TABLE 2

Descriptive Statistics of Trust and Distrust Items in the In-Service Teacher Data Set

	Variable	<i>M</i>	<i>SD</i>	Skewness	Kurtosis	Maximum absolute z-value
Reasons for trust in researchers in general	Expertise	4.20	0.64	-0.64	1.65	5.01
	Integrity	4.06	0.78	-0.91	1.48	3.93
	Benevolence	3.10	0.96	-0.15	-0.53	2.18
Reasons for distrust in researchers in general	Expertise	2.44	0.80	0.41	0.22	3.21
	Integrity	3.20	0.97	-0.24	-0.56	2.28
	Benevolence	4.07	0.82	-0.86	0.76	3.73
Reasons for trust in educational researchers	Expertise	3.63	0.89	-0.80	0.53	2.97
	Integrity	3.66	0.90	-0.93	0.97	2.96
	Benevolence	3.38	0.98	-0.60	-0.24	2.42
Reasons for distrust in educational researchers	Expertise	2.82	0.94	0.28	-0.39	2.31
	Integrity	3.51	1.00	-0.48	-0.46	2.51
	Benevolence	3.32	1.14	-0.18	-0.93	2.03

In the present study, we hypothesized, for example, that expertise-related reasons would be more strongly endorsed by teachers than benevolence-related and integrity-related reasons (H_1 ; see The Present Study section), which can be written as $H_1: \mu_{exp} > (\mu_{int}, \mu_{ben})$. To gain robust relative evidence for or against this hypothesis, we used the following preregistered decision procedure (Rosman & Merk, 2020): First, we computed approximate adjusted fractional Bayes factors (Gu et al., 2019; Hoijtink et al., 2019) for a Bayesian repeated measures ANOVA regarding the comparison of H_1 with the more restricted hypothesis $H_{1R}: \mu_{exp} > \mu_{int} = \mu_{ben}$. This hypothesis states that, on average, higher scores are found for teachers on expertise-related reasons compared with benevolence- and integrity-related reasons, and, furthermore, that the scores of benevolence- and integrity-related reasons are the same. Second, we computed Bayes factors for the comparison of H_1 with the null hypothesis H_0 , which states that all means are equal ($H_0: \mu_{exp} = \mu_{int} = \mu_{ben}$). If the Bayes factors provided evidence for H_1 in both cases, we finally computed a Bayes factor comparing H_1 with its complement H_{1C} . This complement includes any ordering of the means $\mu_{exp}, \mu_{int}, \mu_{ben}$ which does *not* satisfy H_1 (and hence includes, for example, H_0 and H_{1R} , but also mean configurations such as $\mu_{exp} < \mu_{int} < \mu_{ben}$). We intended to label our results as “evidence for H_1 in comparison to H_{1R}, H_0 and H_{1C} ” if, and only if, all three of these Bayes factors provide evidence for H_1 —in all other cases, we intended to use the label “inconclusive.”

This decision procedure, which is exemplarily described here for H_1 , was also used for H_2, H_3, H_4, H_5 , and H_6 . Within each step of the decision procedure, we specified Bayes factors greater as 1/3 but smaller than 3 as rather inconclusive (despite the fact that we will try to refrain from a “dichotomous” interpretation of the resulting Bayes factors).

Technically, the bain framework follows a tradition set by O’Hagan (1995) and uses a fraction of the information in the

data to set the variance of the (normal) prior distribution,

which, in the context of ANOVA, is $\mu_g \sim N\left(\mu_B, \frac{1}{b_g} \cdot \frac{\hat{\sigma}^2}{N_g}\right)$.

This means that the prior distribution of the mean of group g (μ_g) is located at the boundary of the hypotheses under consideration μ_B , and its variance is defined by the residual variance of the ANOVA ($\hat{\sigma}^2$), the size of the group (N_g), and the fraction of the information regarding μ_g in the data $\left(\frac{1}{b_g}\right)$.

Despite the fact that the idea of using a fraction of the data as

minimal training sample for the specification of the prior distribution is well established in the literature (Hoijtink et al., 2019), deciding *which* fraction remains somewhat

subjective. Many authors suggest $b_g = \frac{J}{G} \cdot \frac{1}{N_g}$, whereby G

denotes the number of groups and J the number of con-

straints of H_0 (see Gu et al., 2018, for further elaborations).

To ensure the robustness of our results, we repeated each analysis with $2 \cdot b_g$ and $3 \cdot b_g$, a procedure called “sensitivity analysis,” which is recommended by several checklists for Bayesian analyses (Depaoli & van de Schoot, 2017; van Doorn et al., 2019).

Results

After carefully checking our data with regard to the exclusion criteria specified in our preregistration (i.e., major protocol deviations), we decided not to exclude any cases and proceeded with the data analysis using the full data set. Table 2 includes descriptive statistics on the study variables from the in-service teacher sample. As model assumptions were not heavily violated (see Table 2; Bosman, 2018;

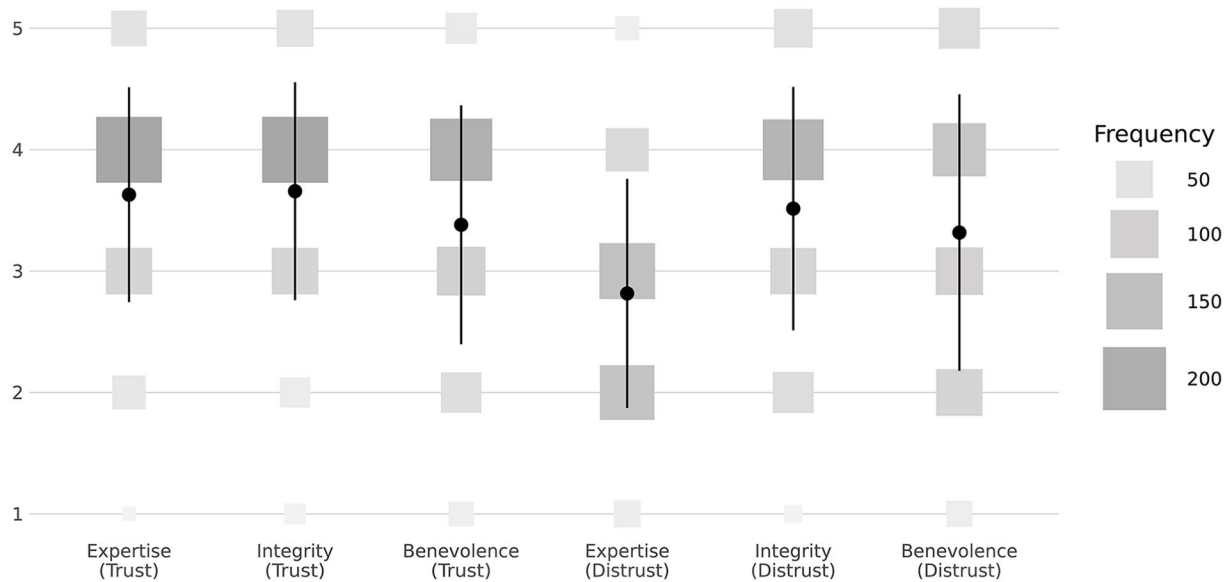


FIGURE 1. *Reasons for trust and distrust in educational researchers: Product plot of the raw data and means $\pm 1*SD$.*

Tabachnick & Fidell, 2014; van Rossum et al., 2013), we used standard (instead of robust) estimators for the mean parameters and their covariance matrices.

Hypotheses 1 and 2

Hypothesis 1 predicts that teachers would score higher on expertise-related reasons compared with benevolence-related reasons and integrity-related reasons when asked about their reasons for trusting educational researchers. Moreover, concerning distrust in educational researchers, Hypothesis 2 suggests that teachers would score lower on expertise-related reasons compared with benevolence-related reasons and integrity-related reasons. Figure 1 and Table 2 provide an overview of these results and not only indicate a confirmation of Hypothesis 2 but surprisingly similar mean values for expertise and integrity as reasons for trust in education researchers (thereby contradicting H_1). In line with this, the Bayes factors (regarding reasons for trust) favored $H_1: \mu_{\text{exp}} > (\mu_{\text{int}}, \mu_{\text{ben}})$ over $H_0: \mu_{\text{exp}} = \mu_{\text{int}} = \mu_{\text{ben}}$ and $H_{1R}: \mu_{\text{exp}} > \mu_{\text{int}} = \mu_{\text{ben}}$, but not against its complement H_{1C} (see Rosman & Merk, 2021a, for the code and Rosman & Merk, 2021b, for the corresponding data and markdown file). According to our preregistered decision procedure, this is an inconclusive result. However, considering our rather conservative hypothesis formulation in the preregistration and the strong descriptive differences between expertise- and benevolence-related reasons, we decided to further investigate this inconclusive result and set up two new exploratory hypotheses using the bain framework: $H_{1a} (\mu_{\text{exp}} > \mu_{\text{int}})$ and $H_{1b} (\mu_{\text{exp}} > \mu_{\text{ben}})$. These additional (thus not

preregistered and to be interpreted with caution) analyses resulted in strong evidence for $\mu_{\text{exp}} = \mu_{\text{int}}$ (against $\mu_{\text{exp}} > \mu_{\text{int}}$; $BF = 30.2$) and $\mu_{\text{exp}} > \mu_{\text{ben}}$ (against $\mu_{\text{exp}} = \mu_{\text{ben}}$; $BF = 2,380$). Hypothesis 2, in contrast, was clearly supported according to the preregistered decision procedure (all $BFs > 50$). All effect sizes can be viewed in Table 3.

Hypotheses 3 and 4

Regarding the reasons for trust and distrust, Hypotheses 3 and 4 predict the same patterns as Hypotheses 1 and 2, but focus on how teachers view researchers in general instead of considering the specific group of educational researchers. Descriptive results of Hypotheses 3 and 4 are depicted in Figure 2 and Table 2. Descriptively, all effects are in line with the hypotheses and are, according to Cohen’s (1988) benchmarks, “small” to “strong” in magnitude (see Table 3). All derived Bayes factors for the decision procedure specified above provided very strong evidence for the preregistered hypotheses (all $BFs > 1,000$; see Rosman & Merk, 2021b).

As the investigation of Hypothesis 1 remained inconclusive, as outlined in the preregistration, we did not explore differences between reasons for trusting educational researchers and researchers in general. To nevertheless gain an insight into this exploratory research question and given that Hypotheses 2 and 4 were both significant, we subsequently conducted these exploratory analyses with regard to the distrust variables. To do so, we tested the following hypotheses against each other and their complements (whereby μ indicates the group mean, “sig” denotes “science/research in general,” and “es” denotes “educational science/research”):

TABLE 3

Effect Sizes (Vargha and Delaney's A; Vargha & Delaney, 2000)

	In-service teachers								Matched general population			
	Educational science				Science (global)				Science (global)			
	Trust		Distrust		Trust		Distrust		Trust		Distrust	
	Int	Ben	Int	Ben	Int	Ben	Int	Ben	Int	Ben	Int	Ben
Exp	.49	.57	.31	.37	.54	.81	.28	.10	.58	.83	.23	.09

Note. Exp = Expertise; Int = Integrity; Ben = Benevolence. The following correspondences between Cohen's d and A may help to interpret the magnitude of the effects: $d = .20 \hat{=} A = .56$, $d = .50 \hat{=} A = .64$, $d = .80 \hat{=} A = .71$; $d = -.20 \hat{=} A = .44$, $d = -.50 \hat{=} A = .36$, $d = -.80 \hat{=} A = .29$.

$$\begin{array}{l}
|\mu_{\text{exp_sig}} - \mu_{\text{int_sig}}| > |\mu_{\text{exp_es}} - \mu_{\text{int_es}}| \ \& \ |\mu_{\text{exp_sig}} - \mu_{\text{ben_sig}}| > |\mu_{\text{exp_es}} - \mu_{\text{ben_es}}| \\
|\mu_{\text{exp_sig}} - \mu_{\text{int_sig}}| = |\mu_{\text{exp_es}} - \mu_{\text{int_es}}| \ \& \ |\mu_{\text{exp_sig}} - \mu_{\text{ben_sig}}| = |\mu_{\text{exp_es}} - \mu_{\text{ben_es}}| \\
|\mu_{\text{exp_sig}} - \mu_{\text{int_sig}}| < |\mu_{\text{exp_es}} - \mu_{\text{int_es}}| \ \& \ |\mu_{\text{exp_sig}} - \mu_{\text{ben_sig}}| < |\mu_{\text{exp_es}} - \mu_{\text{ben_es}}|
\end{array}$$

In this analysis, the first hypothesis was strongly favored by the Bayes factors (against the other two and against its complement; all $BFs > 6,000$). This leads to the conclusion that there is strong evidence for the hypothesis of a relatively stronger smart but evil stereotype regarding researchers in general.

Hypotheses 5 and 6

Hypotheses 5 and 6 focus on how teachers and an age- and SES-matched general population sample compare in their trust and distrust of researchers in general. These comparisons were made using data from the 2018 Science Barometer, a study conducted in August 2018 by means of computer-assisted telephone interviewing. According to Kantar Emnid, who conducted the data collection, the sample is representative for the German general population aged 14 years and older (Wissenschaft im Dialog/Kantar Emnid, 2018). In total, it comprises $N = 1,008$ participants (51% female; 11% 14–19 years; 10% 20–29 years; 14% 30–39 years; 16% 40–49 years; 18% 50–59 years; 13% 60–69 years; 18% 70 years+; 83% former West Germany).

Before testing Hypotheses 5 and 6, we matched participants of our study to the 2018 Science Barometer participants. We thereby used genetic matching (Diamond & Sekhon, 2013; Sekhon, 2011) with replacement to achieve a general population sample with comparable joint distributions concerning age, education, and interest in politics, sports, and science. As can be seen in Table 3, this matching procedure resulted in very similar samples concerning central tendency (maximum difference in means corresponds to an absolute Cohen's d of 0.02) and shape of the distributions (maximum of differences in empirical cumulative distribution functions = .09).

With the matched data set, we conceptually carried out the same analyses as in Hypotheses 3 and 4, but using an estimation method that takes the sample weights (stemming from the matching with replacement) into account. The

corresponding descriptive statistics are illustrated in Figure 3. Our analyses resulted in Bayes factors favoring the hypothesis of higher expertise-related compared with benevolence- and integrity-related reasons for trust (all $BFs > 1,000$), and lower expertise-related compared with benevolence- and integrity-related reasons for distrust (all $BFs > 1,000$). Hypotheses 5 and 6 are thus supported.

The exploratory part of these analyses (see above) was tested by contrasting the reasons for trusting researchers in general between the teacher sample and the matched sample. This was done by comparing Bayes factors for the following hypotheses (whereby μ indicates the group mean and “sig” indicates “scientists/researchers in general”):

$$\begin{array}{l}
\left(\mu_{\text{exp_sig}^{\text{teachers}}} - \mu_{\text{ben_sig}^{\text{teachers}}}\right) > \left(\mu_{\text{exp_sig}^{\text{gen.pop.}}} - \mu_{\text{ben_sig}^{\text{gen.pop.}}}\right) \ \& \\
\left(\mu_{\text{exp_sig}^{\text{teachers}}} - \mu_{\text{int_sig}^{\text{teachers}}}\right) > \left(\mu_{\text{exp_sig}^{\text{gen.pop.}}} - \mu_{\text{int_sig}^{\text{gen.pop.}}}\right) \\
\left(\mu_{\text{exp_sig}^{\text{teachers}}} - \mu_{\text{ben_sig}^{\text{teachers}}}\right) = \left(\mu_{\text{exp_sig}^{\text{gen.pop.}}} - \mu_{\text{ben_sig}^{\text{gen.pop.}}}\right) \ \& \\
\left(\mu_{\text{exp_sig}^{\text{teachers}}} - \mu_{\text{int_sig}^{\text{teachers}}}\right) = \left(\mu_{\text{exp_sig}^{\text{gen.pop.}}} - \mu_{\text{int_sig}^{\text{gen.pop.}}}\right) \\
\left(\mu_{\text{exp_sig}^{\text{teachers}}} - \mu_{\text{ben_sig}^{\text{teachers}}}\right) < \left(\mu_{\text{exp_sig}^{\text{gen.pop.}}} - \mu_{\text{ben_sig}^{\text{gen.pop.}}}\right) \ \& \\
\left(\mu_{\text{exp_sig}^{\text{teachers}}} - \mu_{\text{int_sig}^{\text{teachers}}}\right) < \left(\mu_{\text{exp_sig}^{\text{gen.pop.}}} - \mu_{\text{int_sig}^{\text{gen.pop.}}}\right)
\end{array}$$

As the comparison of Figures 2 and 3 already suggests, the Bayes factors favored the second hypothesis against the two others and its complement (all $BFs > 60$), which can be interpreted as relative evidence for the equality of the magnitude of the pattern between both groups.

Discussion

The present study aimed to investigate in-service teachers' reasons for epistemic trust in educational researchers

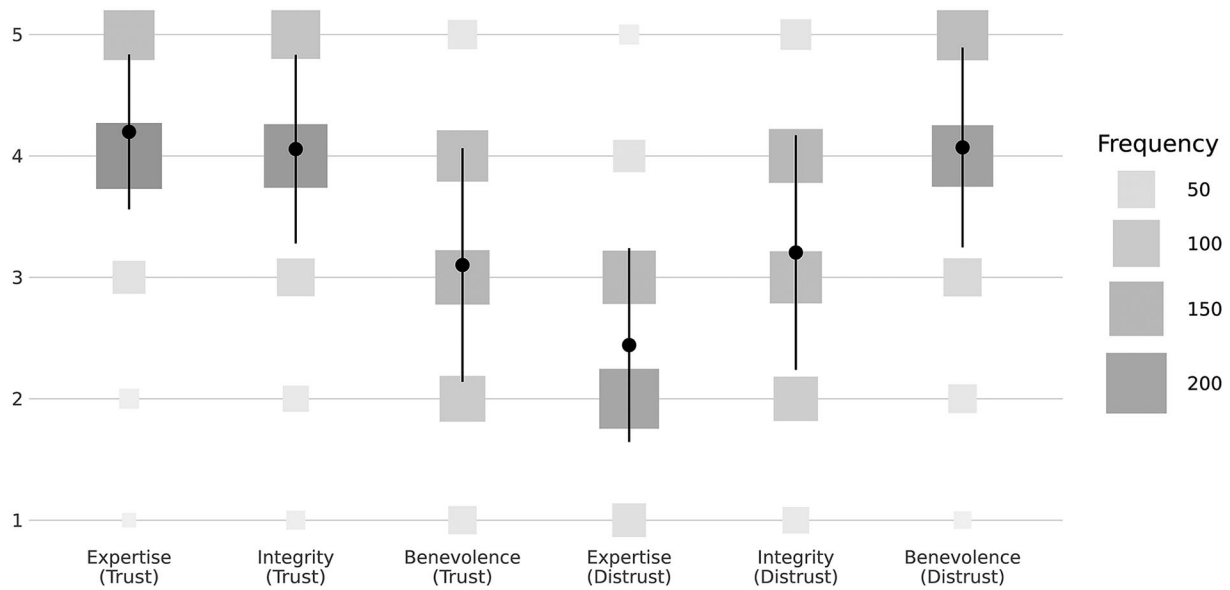


FIGURE 2. *Reasons for trust and distrust in researchers in general: Product plot of the raw data and means $\pm 1*SD$.*

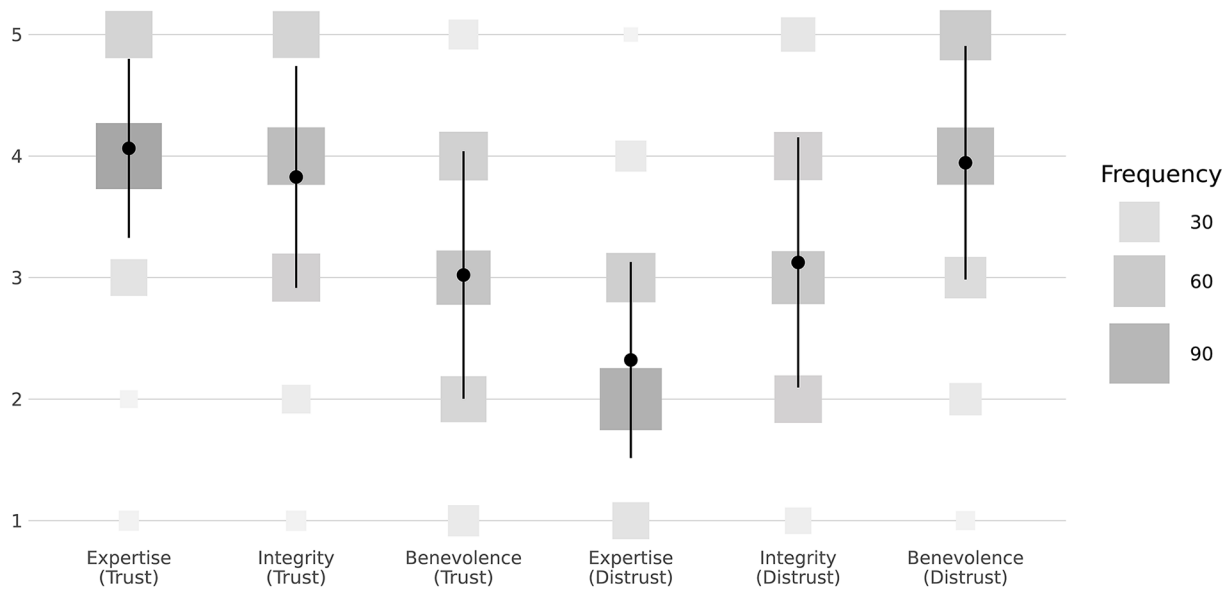


FIGURE 3. *Reasons for trust and distrust in researchers in general: Product plot of the raw data and means $\pm 1*SD$ (matched general population).*

as well as in researchers in general. Previous research has suggested that educational researchers are seen as competent and qualified, but also as having comparably less integrity and benevolence. In line with such results, we expected that our participants would more strongly refer to expertise-related reasons compared with explanations

focusing on benevolence and integrity to justify their trust in (educational) researchers, and that the contrary would be true when justifying their distrust. Furthermore, we expected that this pattern of results would generalize to research in general and that it would also be present in the general population. We formulated three corresponding

sets of preregistered hypotheses (Rosman & Merk, 2020) that were subsequently tested in a sample of 414 German in-service teachers.

Hypotheses 1 and 2

With regard to distrust in educational researchers, our expectations were fully supported: Teachers rated expertise-related reasons for distrust in educational researchers lower than benevolence- and integrity-related reasons. Benevolence and integrity thus seem to play a comparably stronger role in teachers' justifications for distrust in educational researchers. With regard to trust in educational researchers, however, our evidence was rather mixed. First, our confirmatory analyses yielded an inconclusive result, which is why the following interpretations are solely based on exploratory analyses. In these analyses, and in line with our expectations formulated in the first part of Hypothesis 1, participants rated expertise-related reasons for trust in educational researchers as considerably higher compared with benevolence-related reasons. However, we also found evidence that participants rated integrity-related reasons equally high as expertise-related reasons, which clearly contradicts Hypothesis 1. Several explanations for this unexpected finding can be offered. First, a correlation of $r = .39$ between the corresponding expertise items and integrity items (which is somewhat higher than the correlation between benevolence and integrity; see Rosman & Merk, 2021b) indicates that there may be some conceptual overlap between the items assessing expertise and integrity. This is possible because "working according to rules and standards" (i.e., the integrity item; see Table 1) requires knowledge about and thus expertise on such rules and standards in the first place. Therefore, participants may not have differentiated enough between the two items, leading to the almost identical ratings we observed. Furthermore, the notion of "rules and standards" may have been too abstract and ambiguous. In fact, the items from the adjective-based METI have a much stronger affective component (e.g., "dishonest," "insincere," "unfair"; Hendriks et al., 2015) than the simple notion of a low adherence to rules and standards, and this might well explain the discrepancies between our results and the findings by Merk and Rosman (2019). Nonetheless, it should also be noted that we found rather high correlations between our items and the METI items in our pilot study.

Hypotheses 3 and 4

Our second set of hypotheses predicted that the denoted pattern would generalize to research in general. More

specifically, we expected that in-service teachers would, when asked about their reasons for trust in researchers in general, rate expertise-related reasons higher than benevolence- and integrity-related reasons—and that the opposite pattern would come to light when asked about their reasons for distrust in researchers in general. All confirmatory hypotheses were supported; hence, the denoted pattern indeed seems to generalize to higher order domains and is not specific for educational researchers only. Interestingly, these results are in line with the predictions of the TIDE framework (Muis et al., 2006), suggesting that its assumptions may be generalized from epistemic beliefs research to the field of epistemic trust. If we were to speculate, it could thus well be that a more general distrust in researchers (and not only distrust in educational researchers) is responsible for teachers' preference in practically derived knowledge sources over scientific evidence (e.g., Bråten & Ferguson, 2015; Zeuch & Souvignier, 2016). Furthermore, in an additional exploratory analysis, we investigated the magnitude of differences between the smart but evil pattern regarding educational researchers and researchers in general. This yielded strong evidence for a relatively stronger smart but evil stereotype regarding distrust in researchers in general compared with educational researchers. Hence, teachers do not appear to be specifically biased toward educational researchers, or, in other words, the smart but evil pattern might not be specific for educational research, but instead reflect a more general belief pattern that is even stronger in other contexts. However, it should also be pointed out that mean ratings of trust in educational researchers were somewhat lower compared with trust in researchers in general (see Table 1).

Hypotheses 5 and 6

In our third and final set of hypotheses, we investigated whether the general populations' explanations for their trust and distrust in researchers also reflects the smart but evil pattern. We did so by comparing our data with general population data from the 2018 Science Barometer data (Wissenschaft im Dialog/Kantar Emnid, 2018). To draw valid comparisons between both data sets, we matched the Science Barometer data to our sample with regard to age, education, and interest in politics, sports, and science. Again, our confirmatory hypotheses were fully supported—the pattern does not seem to be limited to the educational domain and/or to a sample of in-service teachers. In addition, our exploratory analyses suggested that the magnitude of the smart but evil pattern (regarding research in general) is identical for teachers and for members of the general population. Hence, teachers do not seem to form some kind of special group regarding their epistemic trust; on the contrary, their belief patterns

regarding researchers are largely in line with those of the general population.

Strengths, Limitations, and Directions for Future Research

We first want to point out that even though we used a different theoretical focus, measurement approach, methodology, and sample, our results are largely consistent with Merk and Rosman's (2019) findings. Therefore, one might be tempted to interpret our findings as a confirmation of the smart but evil pattern (Merk & Rosman, 2019). However, we want to emphasize that our items did not measure trust and distrust per se, but instead focused on individual reasons for trusting or distrusting researchers (but note the high correlations between the original and reworded items in our pilot study). Furthermore, in the past few years, several population representative opinion polls have suggested that there is a fair amount of trust in science and scientific practice in general (Pew Research Center, 2019; Wissenschaft im Dialog/Kantar Emnid, 2017). This is also reflected by our data. Even if our pattern of results on the three trust dimensions indicates that the integrity and benevolence of researchers is questioned more often than their expertise, overall mean scores on the respective items were rather high for trust and low for distrust.

On a methodical level, one might also criticize our use of single items. This is especially problematic considering the mixed results found for Hypothesis 1, which might, as outlined above, have been caused by the ambiguous wording of the trust-integrity item. Besides pragmatic reasons (i.e., questionnaire length), we opted for a single-item measurement of trust and distrust because this would allow comparisons with the Science Barometer data (Wissenschaft im Dialog/Kantar Emnid, 2017). This latter issue also relates to the possibility that our results might *not* be caused by the smart but evil pattern, but by specific properties of certain items that might evoke more agreement independently of our participants' underlying perceptions. What speaks against this interpretation, however, is the fact that our expectations regarding both trust *and* distrust were supported, and these were assessed using a substantially different set of items (see Table 1). Furthermore, our pilot study provided evidence for the validity of our items, and it even suggested that they may be well suited to *directly* measure expertise, benevolence, and integrity—in contrast to solely assessing individual reasons for epistemic trust. In sum, these results reduce the possibility that our effects were caused by specific properties of the items. Nevertheless, future research should strive to replicate our findings using the same as well as other methodological approaches, such as interviews or vignette-based experiments. For example, one might design vignettes that describe researchers that are pictured as more or less benevolent, integer, or competent, and subsequently compare these regarding general trust ratings.

It should also be noted that our findings for Hypotheses 3 and 4 are limited to comparisons between science in general and educational research. Future research might investigate whether the smart but evil pattern varies over different types of domains (e.g., well-structured vs. ill-structured disciplines), a question that is receiving increasing attention in the field of epistemic belief research, too (e.g., Rosman et al., 2020; Rowley et al., 2008). With regard to future research on Hypotheses 5 and 6, one might investigate samples across different cultural or socioeconomic contexts, such as collectivist versus individualistic cultures. In fact, our results are limited to a German context, and it might well be that the smart but evil pattern varies over cultures, for example, due to differences in the susceptibility for in-group bias (which might lead, as outlined in the introduction, to a preference of teachers over researchers). In this regard, however, it should be noted that studies on differences in in-group bias across cultures yielded inconsistent results in the past (Brewer & Chen, 2007), which speaks, to a certain extent, against culture as a central determinant of the smart but evil pattern.

Conclusions and Implications for Practice

Taken together, our findings support a central claim recently made by Könneker (2020): At least in the German population, distrust in researchers seems to be mainly based on distrust regarding researchers' intentions and integrity, whereas their perceived expertise may play a less important role. We thereby found this pattern to be rather general and not specifically directed toward educational researchers. Furthermore, teachers' patterns were practically identical to those of the general population. We consider this generalizability across samples and domains as good news for practice, as they suggest that tools and methods aimed at increasing trust in science in general may also tackle teachers' smart but evil patterns regarding educational researchers. On the level of science communication, we (and others, e.g., Könneker, 2018, 2020) see it as crucial that (educational) researchers communicate their findings in an authentic and intelligible manner—not only focusing on fellow researchers but also on other relevant target groups such as preservice and in-service teachers. The emerging trend to complement research articles with so-called plain language summaries (Kerwer et al., 2021; Shailes, 2017) is encouraging in this regard. As outlined by Könneker (2018), it might also help if researchers talked about their motivations for doing research (e.g., curiosity) more openly. Furthermore, transparency and openness are often suggested as key in increasing trust in science (e.g., Vazire, 2017; Wingen et al., 2020). Following these ideas, we would like to emphasize that the adherence to open science standards in the present study not only increases the "hardness" of our evidence—we also sincerely believe that open science practices are suited to increase the value and trustworthiness of science itself in the long term.

Acknowledgments

The publication of this article was funded by the Open Access Fund of the Leibniz Association. We thank Lisa Trierweiler for carefully proofreading the manuscript for spelling and grammar.

ORCID iDs

Tom Rosman  <https://orcid.org/0000-0002-5386-0499>

Samuel Merk  <https://orcid.org/0000-0003-2594-5337>

Notes

1. As a deviation from the preregistration plan, the labeling of the hypotheses has been changed to better fit the preregistered analysis procedure (Hypotheses 1 to 6 instead of H1a, H1b, etc.). Moreover, for reasons of consistency, we replaced the term “scientists” with “researchers” in all hypotheses.

2. Please note that this study was conducted as a result of a reviewer’s suggestion and thus after collecting and analyzing the data from our main study.

3. For reasons of article length, we only present a general overview over the pilot study’s results here—our full analyses can be found in the reproducible documentation of our analyses (Rosman & Merk, 2021a, 2021b).

References

- Bachmann, R., Gillespie, N., & Priem, R. (2015). Repairing trust in organizations and institutions: Toward a conceptual framework. *Organization Studies*, 36(9), 1123–1142. <https://doi.org/10.1177/0170840615599334>
- Bigley, G. A., & Pearce, J. L. (1998). Straining for shared meaning in organization science: Problems of trust and distrust. *Academy of Management Review*, 23(3), 405–421. <https://doi.org/10.2307/259286>
- Bosman, M. L. (2018). *Robust Bayes factors for Bayesian ANOVA: Overcoming adverse effects of non-normality and outliers* [Master thesis, Utrecht University]. <https://dspace.library.uu.nl/handle/1874/370612>
- Bråten, I., & Ferguson, L. E. (2015). Beliefs about sources of knowledge predict motivation for learning in teacher education. *Teaching and Teacher Education*, 50, 13–23. <https://doi.org/10.1016/j.tate.2015.04.003>
- Brewer, M. B., & Chen, Y.-R. (2007). Where (who) are collectives in collectivism? Toward conceptual clarification of individualism and collectivism. *Psychological Review*, 114(1), 133–151. <https://doi.org/10.1037/0033-295X.114.1.133>
- Bromme, R., Thomm, E., & Wolf, V. (2015). From understanding to deference: Laypersons’ and medical students’ views on conflicts within medicine. *International Journal of Science Education, Part B*, 5(1), 68–91. <https://doi.org/10.1080/21548455.2013.849017>
- Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review*, 1(1), 3–25. https://doi.org/10.1207/s15327957pspr0101_2
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). L. Erlbaum Associates.
- Cramer, C. (2013). Beurteilung des bildungswissenschaftlichen Studiums durch Lehramtsstudierende in der ersten Ausbildungsphase im Längsschnitt [Students’ evaluation of educational science courses in the first phase of teacher training: A longitudinal study]. *Zeitschrift Für Pädagogik*, 59(1), 66–82. <https://doi.org/10.3262/ZP1301066>
- Critchley, C. R. (2008). Public opinion and trust in scientists: The role of the research context, and the perceived motivation of stem cell researchers. *Public Understanding of Science*, 17(3), 309–327. <https://doi.org/10.1177/0963662506070162>
- Depaoli, S., & van de Schoot, R. (2017). Improving transparency and replication in Bayesian statistics: The WAMBS-Checklist. *Psychological Methods*, 22(2), 240–261. <https://doi.org/10.1037/met0000065>
- Diamond, A., & Sekhon, J. S. (2013). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics*, 95(3), 932–945. https://doi.org/10.1162/rest_a_00318
- Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika*, 65(2), 241–261. <https://doi.org/10.1007/BF02294377>
- Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878–902. <https://doi.org/10.1037//0022-3514.82.6.878>
- Fiske, S. T., & Dupree, C. (2014). Gaining trust as well as respect in communicating to motivated audiences about science topics. *Proceedings of the National Academy of Sciences of the United States of America*, 111(Suppl. 4), 13593–13597. <https://doi.org/10.1073/pnas.1317505111>
- Gu, X., Hoijtink, H., Mulder, J., & van Lissa, C. J. (2019). *bain: Bayes factors for informative hypotheses*. <https://cran.r-project.org/web/packages/bain/index.html>
- Gu, X., Mulder, J., & Hoijtink, H. (2018). Approximated adjusted fractional Bayes factors: A general method for testing informative hypotheses. *British Journal of Mathematical and Statistical Psychology*, 71(2), 229–261. <https://doi.org/10.1111/bmsp.12110>
- Hendriks, F., Kienhues, D., & Bromme, R. (2015). Measuring laypeople’s trust in experts in a digital age: The Muenster Epistemic Trustworthiness Inventory (METI). *PLOS ONE*, 10(10), Article e0139309. <https://doi.org/10.1371/journal.pone.0139309>
- Hendriks, F., Kienhues, D., & Bromme, R. (2016). Evoking vigilance: Would you (dis)trust a scientist who discusses ethical implications of research in a science blog? *Public Understanding of Science*, 25(8), 992–1008. <https://doi.org/10.1177/0963662516646048>
- Hendriks, F., Seifried, E., & Menz, C. (2021). Unraveling the “smart but evil” stereotype: Pre-service teachers’ evaluations of educational psychology researchers versus teachers as sources of information. *Zeitschrift Für Pädagogische Psychologie*, 35(2–3), 1–15. <https://doi.org/10.1024/1010-0652/a000300>
- Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, 67(1), 88–140. <https://doi.org/10.3102/00346543067001088>
- Hoijtink, H. (2011). *Informative hypotheses: Theory and practice for behavioral and social scientists*. CRC Press.

- Hojtink, H., Mulder, J., van Lissa, C., & Gu, X. (2019). A tutorial on testing hypotheses using the Bayes factor. *Psychological Methods, 24*(5), 539–556. <https://doi.org/10.1037/met0000201>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Hughes, J. N., West, S. G., Kim, H., & Bauer, S. S. (2018). Effect of early grade retention on school completion: A prospective study. *Journal of Educational Psychology, 110*(7), 974–991. <https://doi.org/10.1037/edu0000243>
- Jimmerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review, 30*(3), 420–437. <https://doi.org/10.1080/02796015.2001.12086124>
- Kerwer, M., Chasiotis, A., Stricker, J., Günther, A., & Rosman, T. (2021). Straight from the scientist's mouth: Plain Language Summaries promote laypeople's comprehension and knowledge acquisition when reading about individual research findings in psychology. *Collabra: Psychology, 7*(1), 18898. <https://doi.org/10.1525/collabra.18898>
- Kirschner, P. A. (2017). Stop propagating the learning styles myth. *Computers & Education, 106*, 166–171. <https://doi.org/10.1016/j.compedu.2016.12.006>
- Könneker, C. (2018). *Vertrauen, Misstrauen, Social Media: Schlüsse aus dem Wissenschaftsbarometer 2018* [Trust, mistrust, social media: Conclusions from the 2018 Science Barometer]. <https://www.wissenschaftskommunikation.de/vertrauen-misstrauen-social-media-schluesse-aus-dem-wissenschaftsbarometer-2018-19243>
- Könneker, C. (2020). Wissenschaftskommunikation und Social Media: Neue Akteure, Polarisierung und Vertrauen [Science communication and social media: New actors, polarization, and trust]. In J. Schnurr & A. Mäder (Eds.), *Wissenschaft und Gesellschaft: Ein vertrauensvoller Dialog* [Science and the public: A trustworthy dialogue] (pp. 25–47). Springer. https://doi.org/10.1007/978-3-662-59466-7_3
- Kruschke, J. K. (2010). Bayesian data analysis. *Wiley Interdisciplinary Reviews. Cognitive Science, 1*(5), 658–676. <https://doi.org/10.1002/wcs.72>
- Landrum, A. R., Eaves, B. S., & Shafto, P. (2015). Learning to trust and trusting to learn: A theoretical framework. *Trends in Cognitive Sciences, 19*(3), 109–111. <https://doi.org/10.1016/j.tics.2014.12.007>
- Lewicki, R. J., McAllister, D. J., & Bies, R. J. (1998). Trust and distrust: New relationships and realities. *Academy of Management Review, 23*(3), 438–458. <https://doi.org/10.5465/amr.1998.926620>
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*(3), 709–734. <https://doi.org/10.5465/amr.1995.9508080335>
- Merk, S., & Rosman, T. (2019). Smart but evil? Student-teachers' perception of educational researchers' epistemic trustworthiness. *AERA Open, 5*(3), 233285841986815. <https://doi.org/10.1177/2332858419868158>
- Merk, S., Rosman, T., Muis, K. R., Kelava, A., & Bohl, T. (2018). Topic specific epistemic beliefs: Extending the theory of integrated domains in personal epistemology. *Learning and Instruction, 56*, 84–97. <https://doi.org/10.1016/j.learninstruc.2018.04.008>
- Merk, S., Rosman, T., Rueß, J., Syring, M., & Schneider, J. (2017). Pre-service teachers' perceived value of general pedagogical knowledge for practice: Relations with epistemic beliefs and source beliefs. *PLOS ONE, 12*(9), Article e0184971. <https://doi.org/10.1371/journal.pone.0184971>
- Muis, K. R., Bendixen, L. D., & Haerle, F. C. (2006). Domain-general and domain-specificity in personal epistemology research: Philosophical and empirical reflections in the development of a theoretical framework. *Educational Psychology Review, 18*(1), 3–54. <https://doi.org/10.1007/s10648-006-9003-6>
- Mullen, B., Brown, R., & Smith, C. (1992). Ingroup bias as a function of salience, relevance, and status: An integration. *European Journal of Social Psychology, 22*(2), 103–122. <https://doi.org/10.1002/ejsp.2420220202>
- O'Hagan, A. (1995). Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society, 57*(1), 99–138. <https://doi.org/10.1111/j.2517-6161.1995.tb02017.x>
- Ou, C. X., & Sia, C. L. (2010). Consumer trust and distrust: An issue of website design. *International Journal of Human-Computer Studies, 68*(12), 913–934. <https://doi.org/10.1016/j.ijhcs.2010.08.003>
- Parr, J. M., & Timperley, H. S. (2008). Teachers, schools and using evidence: Considerations of preparedness. *Assessment in Education: Principles, Policy & Practice, 15*(1), 57–71. <https://doi.org/10.1080/09695940701876151>
- Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest, 9*(3), 105–119. <https://doi.org/10.1111/j.1539-6053.2009.01038.x>
- Pew Research Center. (2019). *Trust and mistrust in Americans' views of scientific experts*. https://www.pewresearch.org/science/wp-content/uploads/sites/16/2019/08/PS_08.02.19_trust_in_scientists_FULLREPORT_8.5.19.pdf
- Rosman, T., & Merk, S. (2020). *Preregistration: Teacher's reasons for trust and mistrust in scientific evidence: Reflecting a "smart but evil" stereotype?* PsychArchives. <https://doi.org/10.23668/psycharchives.2691>
- Rosman, T., & Merk, S. (2021a). Code for: Teacher's reasons for trust and distrust in scientific evidence: Reflecting a "smart but evil" pattern? *PsychArchives*. <http://dx.doi.org/10.23668/psycharchives.4892>
- Rosman, T., & Merk, S. (2021b). Data for: Teacher's reasons for trust and distrust in scientific evidence: Reflecting a "smart but evil" pattern? *PsychArchives*. <http://dx.doi.org/10.23668/psycharchives.4891>
- Rosman, T., Seifried, E., & Merk, S. (2020). Combining intra- and interindividual approaches in epistemic beliefs research. *Frontiers in Psychology, 11*, 570. <https://doi.org/10.3389/fpsyg.2020.00570>
- Rowley, M., Hartley, J., Betts, L., & Robinson, E. J. (2008). What makes a research domain more "scientific"? Undergraduate judgements on biology and psychology. *Psychology Learning & Teaching, 7*(2), 16–25. <https://doi.org/10.2304/plat.2008.7.2.16>
- Saunders, M., Dietz, G., & Thornhill, A. (2014). Trust and distrust: Polar opposites, or independent but co-existing? *Human Relations, 67*(6), 639–665. <https://doi.org/10.1177/0018726713500831>

- Saunders, M., & Thornhill, A. (2004). Trust and mistrust in organizations: An exploration using an organizational justice framework. *European Journal of Work and Organizational Psychology, 13*(4), 493–515. <https://doi.org/10.1080/13594320444000182>
- Schnell, R., Hill, P. B., & Esser, E. (2008). *Methoden der empirischen Sozialforschung* [Methods of empirical social research] (8th ed.). Oldenbourg.
- Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review, 25*(1), 128–142. <https://doi.org/10.3758/s13423-017-1230-y>
- Sekhon, J. S. (2011). Multivariate and propensity score matching software with automated balance optimization: The matching package for R. *Journal of Statistical Software, 42*(7). <https://doi.org/10.18637/jss.v042.i07>
- Shailes, S. (2017). Plain-language summaries of research: Something for everyone. *ELife, 6*, e25411. <https://doi.org/10.7554/eLife.25411>
- Sitkin, S. B., & Roth, N. L. (1993). Explaining the limited effectiveness of legalistic “remedies” for trust/distrust. *Organization Science, 4*(3), 367–392. <https://doi.org/10.1287/orsc.4.3.367>
- Sjølie, E. (2014). The role of theory in teacher education: Reconsidered from a student teacher perspective. *Journal of Curriculum Studies, 46*(6), 729–750. <https://doi.org/10.1080/00220272.2013.871754>
- Southerland, S. A., Granger, E. M., Hughes, R., Enderle, P., Ke, F., Roseler, K., Saka, Y., & Tekkumru-Kisa, M. (2016). Essential aspects of science teacher professional development. *AERA Open, 2*(4), 233285841667420. <https://doi.org/10.1177/2332858416674200>
- Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics* (6th ed.). Pearson Education.
- van Doorn, J., van den Bergh, D., Bohm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Raj, A., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E.-J. (2019). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*. Advance online publication. <https://doi.org/10.31234/osf.io/yqxfr>
- van Rossum, M., van de Schoot, R., & Hoijtink, H. (2013). “Is the hypothesis correct” or “is it not.” *Methodology, 9*(1), 13–22. <https://doi.org/10.1027/1614-2241/a000050>
- Vargha, A., & Delaney, H. D. (2000). A critique and improvement of the CL common language effect size statistics of McGraw and Wong. *Journal of Educational and Behavioral Statistics, 25*(2), 101–132. <https://doi.org/10.2307/1165329>
- Vazire, S. (2017). Quality uncertainty erodes trust in science. *Collabra: Psychology, 3*(1), Article 1. <https://doi.org/10.1525/collabra.74>
- Wingen, T., Berkessel, J. B., & Englich, B. (2020). No replication, no trust? How low replicability influences trust in psychology. *Social Psychological and Personality Science, 11*(4), 454–463. <https://doi.org/10.1177/1948550619877412>
- Wissenschaft im Dialog/Kantar Emnid. (2017). *Detaillierte Ergebnisse des Wissenschaftsbarometers 2017 nach Subgruppen* [Detailed results of the 2017 Science Barometer in different subgroups]. https://www.wissenschaft-im-dialog.de/fileadmin/user_upload/Projekte/Wissenschaftsbarometer/Dokumente_17/Wissenschaftsbarometer2017_Tabellenband.pdf
- Wissenschaft im Dialog/Kantar Emnid. (2018). *Detaillierte Ergebnisse des Wissenschaftsbarometers 2018 nach Subgruppen* [Detailed results of the 2018 Science Barometer in different subgroups]. https://www.wissenschaft-im-dialog.de/fileadmin/user_upload/Projekte/Wissenschaftsbarometer/Dokumente_18/Downloads_allgemein/Tabellenband_Wissenschaftsbarometer2018_final.pdf
- Zeuch, N., & Souvignier, E. (2016). Wissenschaftliches Denken bei Lehramts- und Psychologiestudierenden [Scientific thinking in teacher education and psychology students]. In M. Krämer, S. Preiser, & K. Brusdeylins (Eds.), *Berichte aus der Psychologie. Psychologiedidaktik und Evaluation XI [Reports from Psychology: Psychology didactics and evaluation XI]* (1st ed., pp. 175–183). Herzogenrath: Shaker.

Authors

TOM ROSMAN is a senior researcher and head of the research literacy department at ZPID. His research focuses on epistemic beliefs, psychological processes involved in dealing with conflicting information, and teacher education.

SAMUEL MERK is a junior professor for empirical classroom research and school education at PH Karlsruhe, Germany. His research focuses on school education, professional beliefs, and research methods and statistics.