

Measuring Implicit Attitudes with the Inclusion ST-IAT: A Replication and Further Validation

Sergej Wüthrich*

Bern University of Teacher Education, Switzerland

Caroline Sahli Lozano

Bern University of Teacher Education, Switzerland

Assessment of attitudes in inclusive education research is almost exclusively based on self-report scales. This may lead to overestimation of positive attitudes due to social desirability bias, and self-reported attitudes may not capture all relevant aspects of attitudes. Recently, Lüke and Grosche (2018a) proposed a new attitude test based on a single target variant of the Implicit Association Test. In their sample of pre-service teachers, Lüke and Grosche found self-reported attitudes toward inclusion to be related to socially desirable responding. In contrast, implicit attitudes, as measured by the Single Target Implicit Association Test (Inclusion ST-IAT), were unrelated to social desirability bias and neutral overall. Here, we attempted (1) to replicate these findings, and (2) to further test the discriminatory validity of the Inclusion ST-IAT using two samples expected to differ in their attitudes: pre-service teachers with a study program in primary education (PrE) and special education (SpE). In contrast to the findings by Lüke and Grosche (2018a), we found no evidence of a social desirability bias within self-reported attitudes, and implicit attitudes were positive overall. As expected, SpE pre-service teachers had more positive implicit attitudes than PrE pre-service teachers, which speaks to the validity of the Inclusion ST-IAT. We believe the Inclusion ST-IAT, in combination with self-reports, to be a promising tool in assessing attitudes and encourage further research in this field.

Keywords: Implicit Attitudes, Explicit Attitudes, Inclusion, Social Desirability Bias, Pre-Service Teachers, Special Education

INTRODUCTION

An inclusive educational system aims to provide equal educational opportunities and the right to participate in the regular school system to all children. The implementation of inclusive schooling is a challenging process, and teachers' willingness to accept children with special education needs in their classrooms is a necessary premise for success. Attitude research in inclusive education is therefore of high importance, as it can assess the political and ideological climate, identify factors associated with attitudes, monitor changes, and potentially serve as a predictor of inclusive behavior. Not surprisingly, there is ample research on teacher attitudes toward inclusive education (for reviews, see Avramidis & Norwich, 2002; de Boer, Pijl, & Minnaert, 2011; Scruggs & Mastropieri, 1996). However, this research is almost exclusively based on self-reports. Self-reported attitudes often have limited predictive validity of actual behavior because of additional moderating factors such as prevailing social norms or perceived self-efficacy (Ajzen, 1991; Wallace, Paulson, Lord, &

*Please send correspondence to: Sergej Wüthrich, M.Sc., PHBern, Institut für Forschung, Entwicklung und Evaluation, Fabrikstrasse 2a, Bern, Canton of Bern, 3012, Switzerland, Phone: 0041 31 309 22 05, Email: sergej.wuethrich@phbern.ch.

Bond, 2005), which are rarely taken into account in attitude research on inclusive education (Sharma & Mannan, 2015). Another limitation is that the validity of self-reported attitudes critically depends on people's willingness and ability to faithfully and accurately report their attitudes (Bohner & Dickel, 2011). Also, self-reported attitudes are insensitive in capturing automatic associative evaluations, which often guide behavior outside of awareness (Bargh, 1997).

Measuring Implicit Attitudes

To address some of these issues, a variety of implicit attitude measures have been developed in recent years (Nosek, Hawkins, & Frazier, 2011), with the aim to assess attitudes in a more automatic, unconscious way outside of cognitive control. The most often used of these test procedures is the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). The IAT measures associative strengths between concepts by administering a stimulus categorization reaction time task. The rationale behind this method is that strongly associated concepts (e.g., "people without disabilities" and "good" or "people with disabilities" and "bad") will lead to faster categorizations when they require the same response (i.e., the same button press for "people without disabilities" and "good" stimuli) than will weakly associated concepts requiring the same response (i.e., the same button press for "people with disabilities" and "good" stimuli). The IAT has been used for a wide range of topics (e.g., racial prejudice, gender stereotypes, smoking), yielding generally less favorable attitudes when compared to self-report measures. For example, Wilson and Scior (2015) assessed explicit and implicit attitudes toward individuals with intellectual disabilities and found explicit attitudes to be positive overall, but implicit attitudes to be slightly negative. Further, studies have demonstrated that explicit attitudes toward individuals with disabilities were related to socially desirable responding, but implicit attitudes were not (Thomas, Vaughn, Doyle, & Bubb, 2014). Recently, Lüke and Grosche (2018a; 2018b) reported similar findings for attitudes toward inclusive education. They demonstrated that explicit attitudes tend to correlate with socially desirable responding (Lüke & Grosche, 2018a) and are highly modulated by the social context, in that reported attitudes vary depending on the perceived attitudes of the organization conducting the survey (Lüke & Grosche, 2018b). Therefore, assessing attitudes toward inclusion only using self-reports may lead to positively biased conclusions. This may be especially problematic since predictive validity of the IAT on discriminatory behavior has been claimed to outperform explicit self-report measures on normatively sensitive topics (Nosek, Greenwald, & Banaji, 2007; Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013). Furthermore, even if individuals faithfully report positive attitudes, they may still reveal unintentional negative stereotypes and discriminatory behavior in the classroom, even without being aware of it (Hornstra, Denessen, Bakker, van den Bergh, & Voeten, 2010; Markova, Pit-Ten Cate, Krolak-Schwerdt, & Glock, 2016). Children with special education needs are at high risk of suffering from such implicit discrimination because special education needs are often associated with specific labels or educational measures. Such labels, in turn, are associated with overall lower teacher expectations (Greber, Sahli Lozano, & Steiner, 2017) and lower self-perceived social integration of the children (Sahli Lozano, Greber, & Wüthrich, 2017). Implicit teacher beliefs possibly play an important role because such children

are excluded more readily from certain learning activities, more readily receive negative teacher feedback, or are more readily perceived as “difficult” to integrate into social activities.

The Inclusion ST-IAT

Lüke and Grosche (2018a) suggested that the use of implicit measures along with explicit measures in research on attitudes toward inclusive education could provide a more complete picture of prevailing attitudes. They adopted the Single-Target Implicit Association Test (ST-IAT) to assess implicit attitudes toward inclusive education. The ST-IAT is similar to the standard IAT but uses only a single target category instead of two opposing categories (Bluemke & Friese, 2008; Wigboldus, Holland, & van Knippenberg, 2004). In their study, Lüke and Grosche (2018a) surveyed a sample of 163 pre-service teachers. They found that implicit attitudes were generally neutral, in contrast to generally positive explicit attitudes. Explicit and implicit attitudes correlated moderately, demonstrating slightly related but sufficiently distinct constructs. As already noted, explicit attitudes were moderately related to socially desirable responding, but implicit attitudes were not. These results suggest that the Inclusion ST-IAT might be a valuable tool in overcoming a possible social desirability bias in research on attitudes toward inclusive education. Furthermore, because the Inclusion ST-IAT requires automatic affective reactions that are associated with the term “inclusion,” the combination with self-reported attitudes could allow a better understanding of attitude formation, change, and prediction of inclusive behavior.

Importance of Replication and Further Validation of the Inclusion ST-IAT

Although these initial findings are promising, additional research regarding the reliability and validity of the Inclusion ST-IAT is needed. Replication is necessary to determine the reproducibility and validity of findings, and hence is an indispensable key element in science. Awareness of the need for more systematic replication studies has increased in recent years, and concrete recommendations and frameworks for systematic replications have been developed, for example, by Asendorpf et al. (2013) for psychological research in general and by Coyne, Cook, and Therrien (2016) for special education research. Lüke and Grosche (2018a) explicitly encouraged replication of their study and provided all the necessary materials and information to do so. A conceptual replication of their findings – that a social desirability bias leads to an overestimation of positive attitudes and that implicit attitudes are less favorable – with other samples of pre-service teachers - would render their findings more generalizable. Further, replication of a weak but positive relationship between explicit and implicit attitudes would strengthen confidence that the Inclusion ST-IAT truly measures associations related to the concept of inclusion.

In the present study, we therefore attempted to replicate the initial findings by Lüke and Grosche (2018a) and to further validate the Inclusion ST-IAT using the “known groups” method (using two groups theoretically expected to differ in their implicit attitudes; Lane, Banaji, Nosek, & Greenwald, 2007). Specifically, we assessed explicit and implicit attitudes and socially desirable responding in a mixed sample of pre-service teachers with a study program of either primary school education or special education. To consider the replication of the findings by Lüke and Grosche

(2018a) successful, we would expect to find (1) a moderate correlation between implicit and explicit attitudes, (2) a moderate correlation between socially desirable responding and explicit attitudes, and (3) positive explicit attitudes versus non-positive implicit attitudes in the group of PrE pre-service teachers.

To further validate the Inclusion ST-IAT as an appropriate tool for measuring implicit attitudes toward inclusion, the Inclusion ST-IAT should discriminate between groups expected to differ in their implicit attitudes. Attitudes toward inclusive education are generally more positive among special education teachers (e.g., Abegglen, Schwab, & Hessels, 2015; Feyerer, 2014; Kuhl, Redlich, & Schäfer, 2014), which can be attributed – over and above possible self-selection effects such as interest in special education or the motivation to teach in an inclusive environment – to better knowledge, higher sensitivity, and better training with regard to inclusive education (Abegglen et al., 2015; Varcoe & Boyle, 2014). These aspects should also positively relate to implicit attitudes. Indeed, a study by Kessels, Erbring, and Heiermann (2014), which used the traditional IAT method (using “inclusion” and “multi-tracked school system” as opposite categories) found more favorable implicit attitudes among a subsample of SpE pre-service teachers compared to other pre-service teachers. The result of more positive implicit attitudes among SpE pre-service teachers than among PrE pre-service teachers would therefore strengthen confidence in the Inclusion ST-IAT as a valid measurement tool for assessing implicit attitudes toward inclusion.

In contrast to Lüke and Grosche (2018a), we administered both the questionnaires and the ST-IAT in a single online experiment. Hence, positive results would additionally demonstrate the feasibility of economic online assessment compared to laborious on-site lab testing.

METHOD

Sample

The sample consisted of 51 (92% female) PrE pre-service teachers (bachelor's study program in primary school education) and 63 (89% female) SpE pre-service teachers (master's study program in special education) at the Bern University of Teacher Education ($n = 114$ in total). Because the master's study program in special education requires a teacher's diploma, SpE pre-service teachers were generally older ($M = 37.4$, $SD = 10.2$ years) than PrE pre-service teachers ($M = 24.0$, $SD = 8.1$ years). Also, most of them (86%) already had more than three years of teaching experience.

PrE and SpE pre-service teachers were informed about the online experiment during lectures and asked to participate. Participation was voluntary and anonymous. No compensation was given, but participants received online feedback about their results and were fully debriefed about the purpose of the study after completion of the experiment.

One participant (SpE group) was excluded from all analyses related to the Inclusion ST-IAT because of chance performance when classifying the inclusion stimuli in the Inclusion ST-IAT (57.5% correct; > 6 standard deviations below the sample mean). Other than that, no data was excluded from analyses.

Procedure

The experiment was created using the free web-based software PsyToolkit 2.3.6 (Stoet, 2010, 2017). Participants reached the experiment website through a provided link. An automatic browser check excluded mobile phone and tablet users from participating in the study. Participants were informed about the study topic, data collection/data storage, and the technical requirements for participating in the study (a physical keyboard was needed). After giving informed consent to participate in the study, the experiment started with demographic questions (age, gender, study program, teaching experience). Then, a short scale for assessing socially desirable response behavior (Kemper, Beierlein, Bensch, Kovaleva, & Rammstedt, 2012) and a scale for assessing explicit attitudes toward inclusive education (Kunz, Luder, & Moretti, 2010) followed. The experiment ended with the Inclusion ST-IAT (Lüke & Grosche, 2018a) and a rating of the inclusion words used in the ST-IAT (association of the words with the concept inclusion and their emotional valence). Participants spent on average $Md = 19$ min to complete the experiment.

Instruments

The social desirability scale consisted of two subscales (PQ+: overstating positive qualities/NQ-: minimizing negative qualities) with three items each, using a five-point Likert scale. The two subscales have been shown to have sufficient reliability and validity (Kemper et al., 2012). Sum scores for both subscales were calculated (min: 3, max: 15), and scores for the NQ- subscale were reversed so that higher values indicated more socially desirable responding. Note that Lüke and Grosche (2018a) used a different scale (BIDR; Musch, Brockhaus, & Bröder, 2002). In order to motivate as many pre-service teachers as possible to take part in the study voluntarily, we tried to keep the experiment as short as possible. Therefore, we chose to use the much shorter PQ+/NQ-scales instead of the BIDR scale. Unfortunately, the reliability of the scales in our sample was rather poor ($\alpha_{PQ+} = .64$, $\alpha_{NQ-} = .49$), indicating insufficient internal consistency between the items of the two subscales. Accordingly, interpretations regarding the social desirability bias must be taken with caution. The scale for assessing explicit attitudes toward inclusion consisted of 11 items that could be rated on a six-point Likert scale (EZI; Kunz et al., 2010). The same scale was used by Lüke and Grosche (2017). Sum scores for the scale were calculated (min: 11, max: 66).

The Inclusion ST-IAT (Lüke & Grosche, 2018a) combines the procedure of the ST-IAT (Bluemke & Friese, 2008; Wigboldus et al., 2004) with the concept of inclusion and associated terms. We adopted the same procedure used by Lüke and Grosche (2018a; a detailed description can be found there). We used the exact same stimuli (positive, negative, and inclusion words; see Table 1). The number of blocks, their order, and the number of trials were also the same: (1) 20 practice trials for the classification of words with positive or negative valence (training block), (2) 40 practice trials where inclusion words and positive words required the same decision (PI training block), (3) 80 experimental trials with the same categorization scheme (PI block), (4) 40 practice trials where inclusion words and negative words required the same decision (NI training block), and (5) 80 experimental trials with the same categorization scheme (NI block). Also, the individual trial structure was identical (300ms inter-stimulus interval, during which a fixation dot was displayed). Incon-

rect answers were indicated by displaying a red cross. However, one adaption in the procedure was made: if response latency in an individual trial was longer than 3s, a notification was shown to remind the participants to make faster decisions, and the trial was logged as incorrect (this was the case in only 0.3% of all trials). This modification was made to ensure that participants stayed alert during the task. Data preparation and analysis were done the exact way done Lüke and Grosche (2018a): only experimental trials from the PI and NI blocks with correct responses were considered (=95.14% of all trials). The first trial in each block was dropped from analysis. Also, trials with latencies shorter than 300ms were trimmed to 300ms (=0.01% of all trials). Then, the ST-IAT effect *D* for each participant was calculated (Bluemke & Friese, 2008; Lüke & Grosche, 2018a) by subtracting the mean latency in the PI block from the mean latency in the NI block, divided by the standard deviation of latencies in both blocks. Therefore, a positive *D* represents shorter response latencies in the PI block (where “positive” and “inclusion” words were combined) than in the NI block (where “negative” and “inclusion” words were combined) and indicates a positive implicit attitude.

To assess the internal consistency of the ST-IAT, we calculated Cronbach’s α based on the difference in mean response times between PI and NI blocks of 10 trials. Mean response time in the first block of 10 trials in the PI block was subtracted from mean response time in the first block of 10 trials in the NI block. This procedure was repeated for the next seven blocks of 10 trials. Internal consistency was then calculated based on these eight difference scores per participant.

Valence of the stimulus words and association with the concept of inclusion may vary across participants within different study programs. Therefore, we let participants rate valence/association with inclusion of the stimulus words on a seven-point scale at the end of the experiment, as did Lüke and Grosche (2018a). Participants were presented with the inclusion words in a randomized order and rated their valence (anchors: positive–negative); then, the words were presented again, and participants rated their association with the concept inclusion (anchors: associated–not associated).

Table 1. *List of Stimuli Ssed in the Inclusion ST-IAT*

| Positive | Negative | Inclusion |
|--------------------------|--------------------|--|
| Happiness [Glück] | Pain [Schmerz] | School for All [Schule für Alle] |
| Love [Liebe] | Disgust [Ekel] | UN convention [UN-Konvention] |
| Present [Geschenk] | Stink [Gestank] | Individualization [Individualisierung] |
| Recreation [Entspannung] | Ulcer [Geschwür] | Participation [Partizipation] |
| Success [Erfolg] | Threat [Bedrohung] | Disability [Behinderung] |

Note: German stimuli words (in square brackets) were used.

RESULTS

ST-IAT Stimulus Evaluation

Lüke and Grosche (2018a) found the association of the stimulus words with the concept of inclusion to be good and the valence of the stimulus words to be positive in their sample. The same pattern was found here (see Table 2 for descriptive statistics; theoretical mean = 4, lower values indicate better association/more positive valence). Note, however, that the SpE pre-service teachers' ratings indicate generally better associations with the concept of inclusion and a more positive valence of the stimulus words. A nonparametric MANOVA (*R* package *npmv*; Burchett, Ellis, Harrar, & Bathke, 2017) confirmed that there was a significant difference in ratings between SpE and PrE pre-service teachers (association: ANOVA type test $T_A = 6.81$, $p < .01$, mean of relative effects: .61; valence: $T_A = 6.67$, $p < .01$, mean of relative effects: .61).

Table 2. Participant Ratings of the Association and Valence of the “Inclusion” Stimuli

| Stimuli | Association | | Valence | |
|-------------------|----------------------|----------------------|----------------------|----------------------|
| | PrE <i>M (SD)</i> | SpE <i>M (SD)</i> | PrE <i>M (SD)</i> | SpE <i>M (SD)</i> |
| School for All | 1.55 (1.12) | 1.27 (0.63) | 1.55 (0.90) | 1.57 (0.95) |
| Individualization | 2.71 (1.78) | 1.86 (1.38) | 2.02 (1.05) | 1.65 (0.92) |
| Participation | 2.47 (1.71) | 1.57 (2.03) | 2.02 (1.07) | 1.56 (1.00) |
| UN convention | 3.86 (1.66) | 2.92 (1.62) | 3.18 (1.09) | 2.43 (1.24) |
| Disability | 2.69 (1.93) | 2.52 (1.73) | 3.86 (1.25) | 3.35 (1.39) |

Notes. PrE = Primary education pre-service teachers ($n = 51$). SpE = Special education pre-service teachers ($n = 63$). Words were rated on a 7-point scale (values 1-7) for association and valence (anchors: “associated” - “not associated”; “positive” - “negative”). Lower values represent stronger association / more positive valence.

Implicit Attitudes, Explicit Attitudes, and Socially Desirable Responding

In order to compare our results with the findings by Lüke and Grosche (2018a), overall results (disregarding group membership) were considered, with additional subgroup analyses (group of PrE pre-service teachers only) where more appropriate. Descriptive statistics and Cronbach's α of the measurements of interest are presented in Table 3.

Overall, the ST-IAT effect D , which represents implicit attitudes toward inclusion, significantly exceeded zero ($M = 0.23$, $SD = 0.34$): $t(112) = 7.04$, $p < .001$, d

= 0.66. Therefore, and in contrast to Lüke and Grosche, who found implicit attitudes in their sample to be neutral (with $M = 0.03$, $SD = 0.25$, $n = 163$), implicit attitudes were clearly positive in our sample. This was also the case when only considering the more similar group of PrE pre-service teachers: ($M = 0.14$, $SD = 0.35$): $t(50) = 2.9$, $p < .01$, $d = 0.41$.

Scores on the scale measuring explicit attitudes significantly exceeded the theoretical mean of 38.5 ($M = 44.45$, $SD = 7.66$): $t(113) = 8.29$, $p < .001$, $d = 0.78$. Therefore, like Lüke and Grosche ($M = 41.03$, $SD = 8.12$, $n = 161$), we found explicit attitudes to be positive overall in our sample.

Implicit and explicit attitudes were positively correlated, but this correlation did not reach statistical significance: $r = .15$, $n = 113$, $p = .12$, 90% CI [-0.01, 0.30]. Nonetheless, Lüke and Grosche reported a similar positive relationship of $r = .18$ in their sample. Our sample size was smaller, and the absence of statistical significance could be due to a lack of power. Because we used exactly the same instruments, we decided to use Bayesian statistics to assess whether the correlation coefficient found here is actually more consistent with the coefficient found by Lüke and Grosche than with the null hypothesis of no relationship. Bayes factors represent the likelihood-ratio of two alternative hypotheses, given data. They can be useful when non-significant results are obtained because Bayes factors can quantify evidence in favor of a specified alternative hypothesis relative to the null hypothesis (Dienes, 2014). In this case, we compared the relative evidence for the hypothesis that the found coefficient (.15) is of similar size to the coefficient found by Lüke and Grosche (.18) versus the null hypothesis, i.e., that there is actually no relationship. Following the recommendations by Dienes (2014), and after normalization of the Pearson correlation coefficients using Fischer's z transformation, we chose a half-normal prior distribution with a mode of 0 and a standard deviation of 0.18 and calculated the Bayes factor with the online calculator provided by Dienes (n.d.). The resulting Bayes factor was 2.25. This means that given the obtained data, the hypothesis that there is a true correlation coefficient similar to that found by Lüke and Grosche is more than two times more likely than the null hypothesis. Note that this is still anecdotal evidence for the alternative hypothesis, as only Bayes factors above 3 are considered as substantial evidence by conventions in Bayesian statistics (Dienes, 2014). Therefore, more data would be needed to further substantiate this moderate relationship.

Finally, we tested whether socially desirable responding was significantly correlated with explicit attitudes. This was not the case, neither when using the PQ+ subscale ($r_s = .12$, $n = 114$, $p = .19$) nor when using the NQ- subscale ($r_s = .06$, $n = 114$, $p = .53$), nor the whole scale (sum of scores PQ+ and NQ-; $r_s = .06$, $n = 114$, $p = .52$). Note that Spearman correlation was used because of non-normality of the PQ+/NQ- subscale data distribution. In contrast to the findings by Lüke and Grosche, who found a moderate correlation between explicit attitudes and socially desirable responding ($r = .19$); explicit attitudes were not significantly biased by desirable responding in our sample.

Table 3. Descriptive Statistics for the Attitude Measures and Desirable Responding

| | All | | | | PrE | | | | SpE | | | |
|-----------------|----------|-----------|----------|----------|----------|-----------|----------|----------|----------|-----------|----------|----------|
| | <i>M</i> | <i>SD</i> | <i>W</i> | α | <i>M</i> | <i>SD</i> | <i>W</i> | α | <i>M</i> | <i>SD</i> | <i>W</i> | α |
| ST-IAT <i>D</i> | 0.23 | 0.34 | .983 | .78 | 0.14 | 0.35 | .971 | .81 | 0.30 | 0.32 | .982 | .76 |
| EZI | 44.5 | 7.66 | .991 | .80 | 42.7 | 7.59 | .978 | .80 | 45.9 | 7.48 | .983 | .80 |
| PQ+ | 10.2 | 2.04 | .96* | .64 | 10.5 | 2.10 | .964 | .68 | 9.9 | 1.97 | .94* | .63 |
| PQ- | 12.3 | 2.14 | .92* | .49 | 12.4 | 2.12 | .91* | .51 | 12.2 | 2.17 | .92* | .47 |

Notes. All = Whole sample ($N=114$). PrE = Primary education pre-service teachers ($N=51$). SpE = Special education pre-service teachers ($N=63$). ST-IAT *D* = effect of the implicit association test (positive values indicate positive attitude toward inclusion). EZI = score on the explicit attitude scale (theoretical mean: 38.5, higher values indicate positive attitude toward inclusion). PQ+ / PQ- = subscales of the social desirability scale (higher values indicate more socially desirable responding, max score: 15). *W* = Shapiro-Wilk normality test (star denotes significant deviation from normality). α = Cronbach alpha.

Group Differences in Explicit and Implicit Attitudes

In order to assess whether PrE and SpE pre-service teachers differ in their explicit and implicit attitudes toward inclusion, two t-tests were performed for the scores. On the explicit attitude scale, SpE pre-service teachers ($M = 45.9$, $SD = 7.48$) had significantly higher scores than PrE pre-service teachers ($M = 42.7$, $SD = 7.59$): $t(106.4) = 2.24$, $p < .05$, $d = 0.42$. Likewise, mean ST-IAT *D* effect was significantly higher for SpE pre-service teachers ($M = 0.30$, $SD = 0.32$) than for PrE pre-service teachers ($M = 0.14$, $SD = 0.35$): $t(102.5) = 2.38$, $p < .05$, $d = 0.46$. Therefore, SpE pre-service teachers had more favorable explicit and implicit attitudes toward inclusion than PrE pre-service teachers.

Other than that, there were no group differences regarding the two subscales of socially desirable responding (PQ+: $W = 1796$, $p = .27$, $r = -.06$; PQ-: $W = 1659$, $p = .76$, $r = .07$), and groups did not significantly differ in the correlation coefficients of interest (explicit & implicit attitudes: $z_{\text{Difference}} = -1.2$, $p = .23$; explicit attitudes & PQ+: $z_{\text{Difference}} = 1.09$, $p = .28$; explicit attitudes & PQ-: $z_{\text{Difference}} = 0.57$, $p = .57$).

DISCUSSION

One aim of the current study was to replicate previous findings by Lüke and Grosche (2018a). They found the self-reporting of attitudes toward inclusive education to be potentially biased due to socially desirable responding and reported a dissociation of implicit and explicit attitudes. Our results do not directly replicate these findings. However, they do not directly dispute these initial findings either. Apart from this, we were able to validate the discriminatory sensitivity of the Inclusion ST-IAT in using the known group approach. In the following sections, we discuss

our findings with regard to the findings by Lüke and Grosche (2018a), point out possible limitations of our study and the ST-IAT, and close with recommendations and conclusions.

Replication of Previous Findings and Further Validation of the ST-IAT

Whereas Lüke and Grosche (2018a) reported a moderate correlation between explicit attitudes and socially desirable responding, we could not replicate this finding. Considering that the concept of inclusion is highly normative and that there are previous reports of a social desirability bias regarding attitudes toward inclusion (Lüke & Grosche, 2018b) or attitudes toward disability (Thomas et al., 2014), these results are rather unexpected. A reason for this null finding could have been the decision to use a much shorter social desirability scale (the PQ+ and NQ- subscales of the KSE-G, Kemper et al., 2012). The reliability of the scales in our sample was rather poor, which may have obscured a significant relationship with explicit attitudes. Further, the reported correlation by Lüke and Grosche (2018a) was rather small, and the achieved power of our smaller sample to replicate this correlation was only 66%. Therefore, we are not able to draw strong conclusions about the presence or lack of a social desirability bias in our sample. In our view, the possibility of systematic bias when assessing attitudes towards inclusion warrants attention. Although we believe that more research is needed to determine whether, or under which circumstances, socially desirable responding has a meaningful impact on the validity of such self-reported attitudes, heightened awareness of a possible bias certainly helps to improve the quality of research in this field.

In our sample, the correlation between explicit and implicit attitudes was not statistically significant. However, the correlation coefficient was only a little smaller than the coefficient reported by Lüke and Grosche, and a subsequent Bayesian analysis favored the hypothesis of a correlation in this range over the null hypothesis of a zero correlation. Therefore, these results are in line with findings by Lüke and Grosche nevertheless, although more data would be needed to substantiate this relationship. Regarding the validity of the Inclusion ST-IAT, we consider it important that there is at least some shared variance of explicit attitudes toward inclusion because this heightens confidence that the ST-IAT truly measures associations with the concept of inclusion.

In contrast to the results reported by Lüke and Grosche, implicit attitudes in our sample were positive overall, even when considering only the subgroup of PrE pre-service teachers. One reason might be the fact that our sample also reported more positive explicit attitudes toward inclusion than the sample of Lüke and Grosche, indicating more positive attitudes in general. Furthermore, the valence of the inclusion stimuli in the ST-IAT were also rated slightly more positively by our sample, which may have been the reason for more favorable results in the ST-IAT. We believe that stimuli ratings of valence and association with the concept of inclusion are important aspects to consider when using the Inclusion ST-IAT. But all in all, our data do not support the concern that overall implicit attitudes toward inclusion greatly diverge from self-reported attitudes.

Apart from this replication attempt, we aimed to further validate the Inclusion ST-IAT. Based on the rationale that a valid test for implicit attitudes should

reliably discriminate between groups expected to differ in their attitudes, we compared the results of PrE versus SpE student teachers. The results confirmed these expectations: SpE pre-service teachers had more favorable implicit attitudes than PrE pre-service teachers. This adds evidence to the validity of the Inclusion ST-IAT and corroborates previous findings reported by Kessels, Erbring, and Heiermann (2014), who assessed implicit associations toward inclusion with a traditional IAT. However, because Kessels and colleagues (2014) assessed attitudes relative to a multi-tracked system, conclusions regarding “absolute” positive implicit attitudes in their sample cannot be made.

Possible Limitations of our Study and the Inclusion ST-IAT

We tried to adopt the study protocol by Lücke and Grosche (2018a) as closely as possible to ensure reproducibility. However, there were some significant modifications to the study protocol. We collected participant data in a single session during an online experiment. In contrast, assessments of explicit and implicit attitudes by Lücke and Grosche were administered separately, with an interval of one week between the two sessions. It is possible that task order effects influenced our results, i.e., that the assessment of explicit attitudes right before the ST-IAT made attitudes more salient and influenced the implicit attitude measurement. However, it has been demonstrated that assessment order has only a limited effect on results of the IAT (Nosek, Greenwald, & Banaji, 2005). Apart from this, selecting a shorter social desirability scale was rather problematic in that our data suggested insufficient reliability of the instrument.

With regard to our validation attempt, we need to emphasize that our sample consisted of two quite heterogeneous subsamples for which we predicted differences in implicit attitudes. SpE pre-service teachers were generally much older than PrE pre-service teachers, and most of them already had more than three years of teaching experience. It is possible that other factors unrelated to inclusive education (e.g., age; Nosek, Banaji, & Greenwald, 2002) may have contributed to SpE pre-service teachers’ more favorable ST-IAT scores. Further verifying the discriminatory sensitivity of the Inclusion ST-IAT for more similar subsamples (e.g., comparing PrE pre-service teachers who received training in inclusive education or not, or comparing special education teachers working in inclusive settings or special schools) would be a possibility to bolster these findings.

We also suggest randomizing block order for future use of the Inclusion ST-IAT. As far as we correctly interpreted their procedure description, Lücke and Grosche held category combination order constant. Inclusion stimuli were always paired with the positive category before they were paired with the negative category. There is evidence suggesting that this results in a slight positive bias (Greenwald & Nosek, 2001), which could be problematic when comparing ST-IAT effects of different samples tested with different procedures.

In our view, there are also limitations of the Inclusion ST-IAT in its current form. Category labels of an implicit association test should be easily understood by the respondents (Nosek et al., 2005). We believe this is a potential problem because the concept of “inclusion” might be differently interpreted with varying familiar-

ity and knowledge of the subject and local political policies. This potential problem might be solved by providing a clear definition of inclusion before administering the ST-IAT (which we did not provide in our study). Other possible problems might relate to the exemplars used to represent the category. Used stimuli should accurately represent and be highly associated with the category (Nosek et al., 2007). Our results regarding the stimulus evaluation (Table 2) show that association with inclusion varied between stimuli, and more problematically, between subgroups. If stimuli are only loosely associated with a category, category representation might be only weakly activated by stimulus presentation, and there is a risk that participants resort to other stimuli features (e.g., word length) to categorize words, threatening validity of the ST-IAT. Future studies could more systematically explore variance of Inclusion ST-IAT effects when using different sets of stimuli representing inclusion, as this can affect implicit attitude outcome measurements (Nosek et al., 2005). Finally, and most importantly, predictive validity of the Inclusion ST-IAT for “inclusive behavior” has yet to be demonstrated. While there is evidence for predictive validity of implicit association tests toward discriminatory behavior toward minority groups (Greenwald, Poehlman, Uhlmann, & Banaji, 2009), it remains an open question whether more favorable implicit associations with the concept of inclusion directly relate to actual behavior in the classroom.

CONCLUSION

Overall, we interpret our results as encouraging. Although we did not find evidence of socially desirable responding in our sample and implicit attitudes were generally positive, we found a similar correlation between explicit and implicit attitudes to that reported by Lüke and Grosche (2018a) and more positive implicit attitudes among the sample of SpE pre-service teachers, which speaks to the validity of the Inclusion ST-IAT as a measurement tool for implicit attitudes.

In the present study, we demonstrated the feasibility of assessing implicit attitudes online with little time and effort using free, open-source software (Psytoolkit; Stoet, 2017), and we would like to encourage further research in this field. In combination with self-reports, we believe the ST-IAT to be a promising additional tool in research on attitudes toward inclusive education and more studies are needed to further test this potential. Because implicit and explicit self-reported attitudes correlate only moderately, they seem to measure distinct aspects of attitudes. The assessment of implicit attitudes is more robust to social desirability bias and could predict “inclusive behavior” over and above explicit self-reports. Even if individuals faithfully report positive attitudes toward inclusion, they may still reveal unintentional negative stereotypes and discriminatory behavior in an “inclusive” classroom – e.g., regarding gender (Nosek et al., 2009), ethnic background (Markova et al., 2016), or learning disabilities (Hornstra et al., 2010). We believe that considering implicit attitudes in inclusive education research opens new avenues and possibilities in the process of promoting equal access and social justice within the education system.

REFERENCES

- Abegglen, H., Schwab, S., & Hessels, M. G. P. (2015). Einstellung zur Integration: Ergebnisse aus einer Schweizer Onlinestudie mit Lehrkräften und Studierenden [Attitudes towards inclusion: Results from a Swiss online-study with teachers and colleague students]. *Heilpädagogische Forschung, 41*, 184–192.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*, 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J. A., Fiedler, K., ... Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology: Recommendations for increasing replicability. *European Journal of Personality, 27*, 108–119. <https://doi.org/10.1002/per.1919>
- Avramidis, E., & Norwich, B. (2002). Teachers' attitudes towards integration / inclusion: a review of the literature. *European Journal of Special Needs Education, 17*, 129–147. <https://doi.org/10.1080/08856250210129056>
- Bargh, J. A. (1997). The automaticity of everyday life. In *The automaticity of everyday life: Advances in social cognition, Vol. 10*. (pp. 1–61). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Bluemke, M., & Friese, M. (2008). Reliability and validity of the Single-Target IAT (ST-IAT): assessing automatic affect towards multiple attitude objects. *European Journal of Social Psychology, 38*, 977–997. <https://doi.org/10.1002/ejsp.487>
- Bohner, G., & Dickel, N. (2011). Attitudes and attitude change. *Annual Review of Psychology, 62*, 391–417. <https://doi.org/10.1146/annurev.psych.121208.131609>
- Burchett, W. W., Ellis, A. R., Harrar, S. W., & Bathke, A. C. (2017). Nonparametric inference for multivariate data: The R package **npmv**. *Journal of Statistical Software, 76*, 1–18. <https://doi.org/10.18637/jss.v076.i04>
- Coyne, M. D., Cook, B. G., & Therrien, W. J. (2016). Recommendations for replication research in special education: A framework of systematic, conceptual replications. *Remedial and Special Education, 37*, 244–253. <https://doi.org/10.1177/0741932516648463>
- de Boer, A., Pijl, S. J., & Minnaert, A. (2011). Regular primary schoolteachers' attitudes towards inclusive education: a review of the literature. *International Journal of Inclusive Education, 15*, 331–353. <https://doi.org/10.1080/13603110903030089>
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology, 5*, 1–17. <https://doi.org/10.3389/fpsyg.2014.00781>
- Dienes, Z. (n.d.). Bayes. Retrieved October 25, 2017, from http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/Bayes.htm
- Feyerer, E. (2014). Einstellungen und Haltungen zur inklusiven Schule [Attitudes and positions towards inclusion]. *Erziehung und Unterricht, 164*, 219–227.
- Greber, L., Sahli Lozano, C., & Steiner, F. (2017). Lehrpersoneneinschätzungen von Kindern mit integrativen schulischen Massnahmen [Teachers' perceptions of children with and without integrative programs]. *Empirische Pädagogik, 31*, 303–322.
- Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology, 108*, 553–561. <https://doi.org/10.1037/pspa0000016>
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology, 74*, 1464–1480.
- Greenwald, A. G., & Nosek, B. A. (2001). Health of the Implicit Association Test at Age 3. *Experimental Psychology, 48*, 85–93. <https://doi.org/10.1026//0949-3946.48.2.85>
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*, 17–41. <https://doi.org/10.1037/a0015575>

- Hornstra, L., Denessen, E., Bakker, J., van den Bergh, L., & Voeten, M. (2010). Teacher attitudes toward dyslexia: Effects on teacher expectations and the academic achievement of students with dyslexia. *Journal of Learning Disabilities*, 43, 515–529. <https://doi.org/10.1177/0022219409355479>
- Kemper, C. J., Beierlein, C., Bensch, D., Kovaleva, A., & Rammstedt, B. (2012). *Eine Kurzsкала zur Erfassung des Gamma-Faktors sozial erwünschten Antwortverhaltens: die Kurzsкала Soziale Erwünschtheit-Gamma (KSE-G) [A short scale for assessing the gamma-factor of social desirable response behavior: The short scale Social Desirability-Gamma (KSE-G)]. (GESIS Working Papers 2012|25)*. Köln: GESIS. Retrieved from <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-339589>
- Kessels, U., Erbring, S., & Heiermann, L. (2014). Implizite Einstellungen von Lehramtsstudierenden zur Inklusion [Student teachers' implicit attitudes towards inclusion]. *Psychologie in Erziehung und Unterricht*, 61, 189–202. <https://doi.org/http://dx.doi.org/10.2378/peu2014.art15d>
- Kuhl, J., Redlich, H., & Schäfer, L. (2014). Einstellungen verschiedener Lehrergruppen gegenüber Menschen mit geistiger Behinderung [Attitudes of different groups of teachers towards people with intellectual disabilities]. *Zeitschrift Für Bildungsforschung*, 4, 271–287. <https://doi.org/10.1007/s35834-014-0109-5>
- Kunz, A., Luder, R., & Moretti, M. (2010). Die Messung von Einstellungen zur Integration (EZI) [Measuring attitudes toward inclusion]. *Empirische Sonderpädagogik*, 83–94.
- Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and using the Implicit Association Test: IV: What we now (so far) about the method. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 59–102). New York, NY: Guilford Press.
- Lüke, T., & Grosche, M. (2018a). Implicitly measuring attitudes towards inclusive education: A new attitude test based on single-target implicit associations. *European Journal of Special Needs Education*, 33, 427–436. <https://doi.org/10.1080/08856257.2017.1334432>
- Lüke, T., & Grosche, M. (2018b). What do I think about inclusive education? It depends on who is asking: Experimental evidence for a social desirability bias in attitudes towards inclusion. *International Journal of Inclusive Education*, 22, 38–53. <https://doi.org/10.1080/13603116.2017.1348548>
- Markova, M., Pit-Ten Cate, I., Krolak-Schwerdt, S., & Glock, S. (2016). Preservice teachers' attitudes toward inclusion and toward students with special educational needs from different ethnic backgrounds. *The Journal of Experimental Education*, 84, 554–578. <https://doi.org/10.1080/00220973.2015.1055317>
- Musch, J., Brockhaus, R., & Bröder, A. (2002). Ein Inventar zur Erfassung von zwei Faktoren sozialer Erwünschtheit. *Diagnostica*, 48, 121–129. <https://doi.org/10.1026//0012-1924.48.3.121>
- Nosek, B. A., Banaji, M., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, 6, 101–115. <https://doi.org/10.1037//1089-2699.6.1.101>
- Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, 31, 166–180. <https://doi.org/10.1177/0146167204271418>
- Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In *Social psychology and the unconscious: The automaticity of higher mental processes*. (pp. 265–292). New York, NY, US: Psychology Press.

- Nosek, B. A., Hawkins, C. B., & Frazier, R. S. (2011). Implicit social cognition: from measures to mechanisms. *Trends in Cognitive Sciences*, 15, 152–159. <https://doi.org/10.1016/j.tics.2011.01.005>
- Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., ... Greenwald, A. G. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences*, 106, 10593–10597. <https://doi.org/10.1073/pnas.0809921106>
- Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, 105, 171–192. <https://doi.org/10.1037/a0032734>
- Sahli Lozano, C., Greber, L., & Wüthrich, S. (2017). Subjektiv wahrgenommenes Integriertsein von Kindern in Schulsystemen mit integrativen Massnahmen [Student's perception of their integration at schools with integrative programs]. *Empirische Pädagogik*, 31, 284–302.
- Scruggs, T. E., & Mastropieri, M. A. (1996). Teacher perceptions of mainstreaming/inclusion, 1958–1995: A research synthesis. *Exceptional Children*, 63, 59–74. <https://doi.org/10.1177/001440299606300106>
- Sharma, U., & Mannan, H. (2015). Do attitudes predict behaviour – An (un)solved mystery? In P. Jones & S. Danforth (Eds.), *Foundations of Inclusive Education Research* (Vol. Volume 6, pp. 115–131). Emerald Group Publishing Limited.
- Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, 42, 1096–1104. <https://doi.org/10.3758/BRM.42.4.1096>
- Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44, 24–31. <https://doi.org/10.1177/0098628316677643>
- Thomas, A., Vaughn, E. D., Doyle, A., & Bubb, R. (2014). Implicit association tests of attitudes toward persons with disabilities. *The Journal of Experimental Education*, 82, 184–204. <https://doi.org/10.1080/00220973.2013.813357>
- Varcoe, L., & Boyle, C. (2014). Pre-service primary teachers' attitudes towards inclusive education. *Educational Psychology*, 34, 323–337. <https://doi.org/10.1080/01443410.2013.785061>
- Wallace, D. S., Paulson, R. M., Lord, C. G., & Bond, C. F., Jr. (2005). Which behaviors do attitudes predict? Meta-analyzing the effects of social pressure and perceived difficulty. *Review of General Psychology*, 9, 214–227. <https://doi.org/10.1037/1089-2680.9.3.214>
- Wigboldus, D. H. J., Holland, R. W., & van Knippenberg, A. (2004). *Single target implicit associations*. Unpublished Manuscript.
- Wilson, M. C., & Scior, K. (2015). Implicit attitudes towards people with intellectual disabilities: Their relationship with explicit attitudes, social distance, emotions and contact. *PLOS ONE*, 10(9), e0137902. <https://doi.org/10.1371/journal.pone.0137902>