Measuring those who have their minds set: An item-level meta-analysis of the implicit theories of intelligence scale in education

Ronny Scherer [*], Diego G. Campos

Centre for Educational Measurement at the University of Oslo (CEMO), Faculty of Educational Sciences, University of Oslo, (CEMO), Norway

A R T I C L E   I N F O

A B S T R A C T

Fixed and growth mindsets represent implicit theories about the nature of one's abilities or traits. The existing body of research on academic achievement and the effectiveness of mindset interventions for student learning largely relies on the premise that fixed and growth mindsets are mutually exclusive. This premise has led to the common practice in which measures of one mindset are reversed and then assumed to represent the other mindset. Focusing on K-12 and university students ($N = 27328$), we tested the validity of this practice via a comprehensive item-level meta-analysis of the Implicit Theories of Intelligence Scale (ITIS). By means of meta-analytic structural equation modeling and network analysis, we examined (a) the ITIS item-item correlations and their heterogeneity across 32 primary studies; (b) the factor structure of the ITIS, including the distinction between fixed and growth mindset; and (c) moderator effects of sample, study, and measurement characteristics. We found positive item-item correlations within the sets of fixed and growth mindset items, with substantial between-study heterogeneity. The ITIS factor structure comprised two moderately correlated mindset factors ($\rho = 0.63$–0.65), even after reversing one mindset scale. This structure was moderated by the educational level and origin of the student sample, the assessment mode, and scale modifications. Overall, we argue that fixed and growth mindsets are not mutually exclusive but correlated constructs. We discuss the implications for the assessment of implicit theories of intelligence in education.

1. Introduction

Much of the research in education focuses on identifying possible factors that may determine students' learning and academic success. Over the last few decades, one of these factors has gained considerable attention, both in small- and large-scale interventions (e.g., Sisk et al., 2018; Yeager et al., 2019) and studies establishing the positive link to academic achievement (e.g., Costa & Faria, 2018; Petscher et al., 2017): students' implicit theories or so-called mindsets (Dweck & Yeager, 2019). Dweck (2000) conceptualized mindsets as self-theories about the nature of one's psychological attributes, such as abilities (e.g., intelligence, mathematical skills, creativity) or traits (e.g., personality, moral orientations, emotions). These self-theories can take two forms: People can hold *entity theories*, believing that their psychological attributes are fixed and can hardly change (*fixed mindset*) or *incremental theories*, believing

---

* Corresponding author. Centre for Educational Measurement at the University of Oslo (CEMO), Faculty of Educational Sciences, University of Oslo, Postbox 1161, Forskningsparken, NO-0318, Oslo, Norway.
  E-mail address: ronny.scherer@cemo.uio.no (R. Scherer).

that their psychological attributes are malleable, controllable, and can develop (*growth mindset*; Dweck & Leggett, 1988). The extant body of research on mindsets is largely based on the premise that fixed and growth mindsets are mutually exclusive and that the absence of one suggests the presence of the other—this implies that a student cannot hold both mindsets at the same time (Lüftenegger & Chen, 2017).

This premise has impacted mindset assessment practices in education to a large extent. Specifically, most empirical studies assessing mindsets via students' self-reports entertained one of the following practices (e.g., Burnette et al., 2013; Costa & Faria, 2018; Lüftenegger & Chen, 2017; OECD, 2021; Yeager & Dweck, 2020): (a) Assessing both fixed and growth mindset with multiple items and reverse-coding one to align it with the other; (b) Assessing either fixed or growth mindset with multiple items and assuming that low scores of one mindset indicate high scores of the other; and (c) Assessing one mindset with a single item under the same assumption as in (b). However, more and more evidence has accumulated that fixed and growth mindsets are correlated yet not mutually exclusive constructs (e.g., Cook et al., 2017; Diseth et al., 2014), and students can indeed hold both beliefs (e.g., Burgoyne & Macnamara, 2021; Petscher et al., 2017).

The present study is aimed at clarifying the consequences of the reversing practice of mindset measures meta-analytically. Utilizing primary study data from the most commonly used mindset assessment, the Implicit Theories of Intelligence Scale (ITIS) developed by Dweck (2000), we perform an item-level meta-analysis to examine the relations among fixed- and growth-mindset items, their heterogeneity, the factor structure of the ITIS, and possible moderators thereof. Ultimately, our methodological review provides recommendations guiding the use of the ITIS to assess students' implicit theories of intelligence in education. Besides, we provide meta-analytic evidence on the existence of a single or two correlated mindsets.

2. Theoretical background

2.1. Implicit theories of intelligence in education

Implicit theories represent a person's beliefs about the nature and workings of a psychological attribute or phenomenon (Dweck et al., 1995; Sternberg, 1985). These theories can vary substantially between two persons, as they rely on, for instance, individual experiences, knowledge, and perceptions of the social world (Schunk, 1995). Taking a socio-cognitive perspective, Carol S. Dweck and others have coined the terms "implicit theories of intelligence" or "mindsets" to describe these self-theories (Dweck & Leggett, 1988; Dweck & Yeager, 2019). A person can hold incremental (growth) or entity (fixed) theories (mindsets) about his or her intellectual abilities or psychological attributes, yet not simultaneously, as the current body of research assumes (Lüftenegger & Chen, 2017).

Implicit theories of intellectual abilities have gained considerable attention in educational research, and at least two lines of research have emerged (Dweck & Yeager, 2019): (a) the effectiveness of interventions promoting students' growth mindsets; and (b) the relations between mindsets and educational achievement. While reviewing these research lines in great detail is beyond the scope of this article, we still point to some of their key results. By and large, mindset interventions targeted at developing students' growth mindsets and, ultimately, improving learning and academic achievement seem effective and scalable, yet with small effects and large heterogeneity (Sisk et al., 2018; Yeager et al., 2019). Some studies have testified that these interventions were especially effective for low-achieving and at-risk students (Claro et al., 2016; Paunesku et al., 2015; Sarrasin et al., 2018). Studies focusing on the relations between growth mindset and academic achievement resulted in positive yet small correlations ($r = 0.07$–0.10), again with substantial heterogeneity (Burnette et al., 2013; Costa & Faria, 2018; Sisk et al., 2018). Burnette et al. (2013) established meta-analytically that the mindset-achievement relationship was partially mediated by students' goal orientations and self-regulation. Moreover, some evidence suggests that the mindset-achievement relations may vary across domains (Costa & Faria, 2018). Similar to self-beliefs about one's abilities (e.g., self-concept, self-efficacy), mindsets may indeed be domain-specific (Hass et al., 2017). For instance, Lewis et al. (2021) showed that domain-specific mindsets and a global mindset co-exist—a finding similar to the hierarchical structure of academic self-concepts (Arens et al., 2020).

Clearly, much of the evidence on the effectiveness of mindset interventions and the mindset-achievement relationship abounds in heterogeneous findings with varying effect sizes between studies. Several meta-analyses and large-scale studies explained some of this heterogeneity. For instance, Costa and Faria (2018) found higher correlations between growth mindset and academic achievement for middle-school students ($r = .15$) than for high-school and college students ($r = 0.06$–0.09). A similar moderation effect across developmental stages was backed by Sisk et al. (2018). These stages are, among others, characterized by different stages of cognition, performance on intelligence tests, and structures of the respective measures (Schroeders et al., 2015). Besides, the cultural background of the student samples plays another key role in explaining heterogeneity: Costa and Faria (2018) reported moderator effects of cultural background on the growth mindset-achievement relationship with positive and significant correlations for samples from Asia and Oceania, yet insignificant correlations for samples from Europe and North America. Moreover, the fixed mindset-achievement correlations were negative and statistically significant for North American samples, positive for European samples, and insignificant for Asian samples. The Programme for International Student Assessment (PISA) administered a single item measuring students' fixed mindset in 2018 and examined its relation to reading achievement. Next to the variation in this relation across countries, one finding stood out: "In East Asian countries, growth mindset was not as strongly associated with academic performance as in most OECD countries" (OECD, 2021, p. 17). Yeager and Dweck (2020) discussed this observation and argued that cross-cultural heterogeneity was likely to exist in any mindset study, and that researchers should consider exploring it.

The remaining evidence on possible explanatory variables is, in our reading, limited and adds only a few measurement characteristics: Costa and Faria (2018) found moderator effects of possible modifications of the ITIS (e.g., item wordings, translations, or response options), especially on the fixed mindset-achievement relation (original ITIS: $r = -0.22$, modified scale: $r = 0.02$). Sisk et al.

(2018) observed differences in the effectiveness of growth-mindset interventions between computer-based ($d = 0.03$) and other formats ($d = 0.06$–0.27). Overall, our brief review revealed that sample and measurement characteristics can explain heterogeneity to some extent.

2.2. Measurement issues

Mindset assessments are largely based on students' self-reports, and the Implicit Theories of Intelligence Scale (ITIS) is one the most popular, if not the most popular assessment (Lüftenegger & Chen, 2017). The scale has been administered in a plethora of empirical studies and domains (Costa & Faria, 2018), has been translated into several languages (e.g., Wang & Ng, 2012), and has been used as a criterion to validate other mindset assessments (e.g., Burgoyne & Macnamara, 2021). The ITIS captures students' agreement with a set of statements corresponding to incremental and entity theories and comes in two versions, a six-item version for children (age 10 and older) and an eight-item version for adults (see Table 1; Dweck, 2000). While the items measuring entity theories contain mainly negative wordings to express the fixed nature of the underlying mindset (e.g., "You have a certain amount of intelligence, and you really can't do much to change it."), the incremental theory items are positively formulated to express the orientation towards growth (e.g., "No matter who you are, you can change your intelligence a lot."). While discussing the suitability of students' self-reports to capture their implicit theories is beyond the scope of this article, we notice that these assessments have their limitations, such as possible acquiescence bias or inconsistent response patterns (Freund & Kasten, 2012; Steinmann et al., 2021).

As noted earlier, three practices have dominated the assessment of mindsets (Lüftenegger & Chen, 2017): The first practice is to *assess both mindsets*. Specifically, researchers administer the full six- or eight-item version of the ITIS, extract students' responses, and then reverse-code one item set (e.g., the growth mindset items) to align it with the other set. Oftentimes, the sets item responses are then combined into a single scale score (Yeager & Dweck, 2020). This practice heavily relies on the assumption that the two mindsets are mutually exclusive and that students cannot hold both simultaneously. Recently, some evidence occurred that the ITIS, in fact, measures two correlated latent variables that correspond to the two mindsets (e.g., Glerum et al., 2020; Li & Bates, 2020; Lou et al., 2021), thus challenging this assumption. Research on self-report scales with mixed-worded items deals with a similar problem, and the respective body of evidence suggested that item reversing creates multidimensionality in the item response data (e.g., Kam, 2016). In this sense, the reversing practice may not necessarily lead to a valid representation of students' mindsets as a single construct.

Another common practice is to *assess only one* of the two implicit theories via *multiple items*. This practice is also based on the assumption that the presence of the one mindset indicates the lack of the other. While this practice may be cost- and time-efficient, given that only three or four ITIS items are administered, it is likely to be prone to acquiescence bias and bias in the estimation of the growth- or fixed-mindset scores (Lüftenegger & Chen, 2017).

Similarly, the third practice is to *assess one mindset with a single item*. Clearly, such single-item measures are likely be prone to acquiescence bias and measurement error (OECD, 2021). In their study of three- and single-item measures of fixed mindsets, Rammstedt et al. (2021) showed that the relations to other constructs can differ substantially between the two types of measures and that single-item measures suffer from low test-retest reliabilities. This practice also assumes the orthogonality of the two mindsets. To our best knowledge, the extent to which the assumption of mutually exclusive mindsets underlying the reversing practices in fact holds is still unclear and requires empirical backing.

Table 1
Wordings and subscale assignments of the items in the six- and eight-item version of the ITIS.

Original Stimuli
Implicit Theories of Intelligence Scale for Children—Self Form
"Read each sentence below and then circle the one number that shows how much you agree with it. There are no right or wrong answers." (see Dweck, 2000, p. 177*, p. 177)*
Implicit Theories of Intelligence Scale for Adults—Self Form
"This questionnaire has been designed to investigate ideas about intelligence. There are no right or wrong answers. We are interested in your ideas.
Using the scale below, please indicate the extent to which you agree or disagree with each of the following statements by writing the number that corresponds to your opinion in the space next to each statement." (see Dweck, 2000, p. 178*, p. 178)*

Original Response Options
1 = "Strongly agree", 2 = "Agree", 3 = "Mostly agree", 4 = "Mostly disagree",
5 = "Disagree", 6 = "Strongly disagree"

Item	Wording	Subscale
1	You have a certain amount of intelligence, and you really can't do much to change it.	Fixed
2	Your intelligence is something about you that you can't change very much.	Fixed
3	You can learn new things, but you can't really change your basic intelligence.	Fixed
4	No matter who you are, you can change your intelligence a lot.	Growth
5	You can always greatly change how intelligent you are.	Growth
6	No matter how much intelligence you have, you can always change it quite a bit.	Growth
7	To be honest, you can't really change how intelligent you are.	Fixed
8	You can change even your basic intelligence level considerably.	Growth

Note. The six-item ITIS comprises items 1–6. Item wordings correspond to those of Dweck's Implicit Theories of Intelligence Scales for Children (age 10 and older) and for Adults—Self Forms (Dweck, 2000).

2.3. The present study

To summarize, the extant literature on students' implicit theories of intelligence (i.e., mindsets) is largely based on the premise that entity (i.e., fixed mindset) and incremental (i.e., growth mindset) theories are mutually exclusive (Lüftenegger & Chen, 2017). This premise has impacted current assessment practices of mindsets, such as assessing mindsets by (a) item sets that measure both fixed and growth mindsets, one of which is subsequently reversed to align with the other; (b) item sets that measure either fixed or growth mindsets under the assumption that agreement with one set indicates disagreement with the other; and (c) single items (Yeager & Dweck, 2020). Despite the methodological challenges associated with these practices, the substantive assumption that students cannot hold both fixed and growth mindsets at the same time seems questionable (e.g., Glerum et al., 2020; Lüftenegger & Chen, 2017; Tempelaar et al., 2015). Given the heterogeneous evidence on the existence of a single mindset factor or two correlated mindset factors representing entity and incremental theories, Lüftenegger and Chen (2017) called for large-scale evaluations and replications of the evidence base on the relationship between the two implicit theories. Identifying the possible sources of this heterogeneity is also key to understanding how the study, sample, and measurement contexts may impact the evidence base (see also Carpenter et al., 2016).

The present study is primarily aimed at synthesizing the evidence on the relationship between fixed and growth mindsets of intelligence in the context of education. Ultimately, we aim to clarify whether a single mindset or two correlated mindset factors exist. Our secondary goal is to illustrate the potential of item-level meta-analysis and meta-analytic structural equation modeling for examining research questions concerning educational assessment and measurement. Focusing on the ITIS, we specifically address three research questions (RQs):

RQ 1. To what extent are the ITIS items correlated, and how do these correlations vary across study samples? (Pooled correlations and heterogeneity)

RQ 2. To what extent can the single- and two-factor models represent the structure of the ITIS, and what characterizes the best-fitting model? (Factor structure and model parameters)

RQ 3. Which study, sample, and measurement characteristics moderate the model parameters in the final ITIS factor model? (Moderation of model parameters)

3. Methods

3.1. Literature search and screening

To identify the literature eligible for addressing our RQs, we performed systematic searches in the databases PsycINFO, ERIC, PsyArXiv Preprints, and ProQuest Dissertations & Theses in August 2021, using "mindset" OR "implicit theor*" and "assessment" OR "measurement" OR "scale" as key elements of the search terms. We further supplemented these searches by reviewing the references in the comprehensive review by Costa and Faria (2018) and by searching for raw data in the OECD (https://www.oecd.org/pisa/data/)
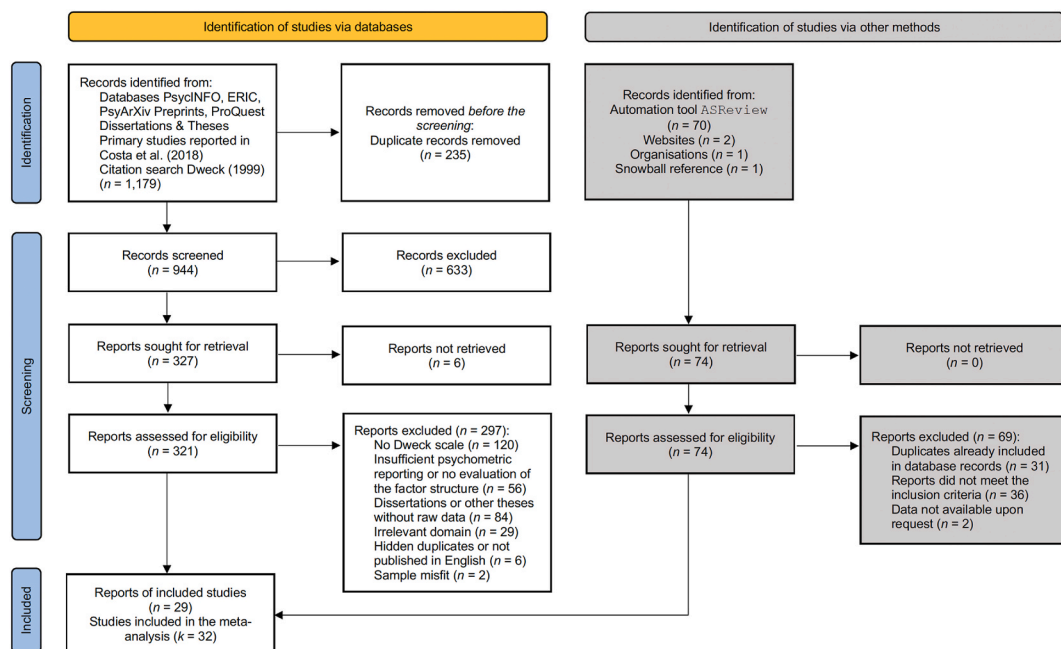

Fig. 1. PRISMA flow diagram of the search, screening, and inclusion processes.

Table 2
Coded features of the primary studies, samples, and measures in the meta-analytic sample.

Feature	Number of study samples (m)	Proportion
Study and sample characteristics		
Publication status		
Published	34	87.2%
Grey literature	5	12.8%
Study purposes		
Associations with other constructs	18	46.1%
Evaluation of an intervention	1	2.6%
Instrument validation	16	41.0%
Longitudinal changes and/or associations	4	10.3%
Study design		
Cross-sectional	32	82.1%
Longitudinal	6	15.4%
(Quasi-)Experimental	1	2.6%
Availability of raw data		
Yes	7	17.9%
No	32	82.1%
Source of item-item correlations		
Raw data	7	17.9%
Observed item-item covariances	1	2.6%
Observed item-item correlations	10	25.6%
Factor loadings (based on EFA or CFA)	21	53.9%
Educational level		
School	24	61.5%
University	15	38.5%
Measurement characteristics		
Response options		
4 options	2	5.1%
5 options	6	15.4%
6 options	24	61.5%
7 options	7	17.9%
Assessment mode		
Paper-and-pencil	24	61.5%
Computer-based and online	15	38.5%
Number of items		
3 items	9	23.1%
4 items	10	25.6%
5 items	6	15.4%
6 items	4	10.3%
8 items	10	25.6%
Mindset domain(s)		
Fixed mindset only	14	35.9%
Growth mindset only	1	2.6%
Fixed and growth mindset	24	61.5%
Scale modifications		
Yes	13	33.3%
No	26	66.7%
Reporting of reliability coefficient(s)		
Yes	33	84.6%
No	6	15.4%
Representation by latent variables		
Yes	25	64.1%
No	14	35.9%
Number of factors		
One factor	21	53.8%
Two factors	17	43.6%
Three factors	1	2.6%
Testing of multiple measurement models		
Yes	23	59.0%
No	16	41.0%
Type of measurement model		
Confirmatory factor analysis (CFA)	30	76.9%
Exploratory factor analysis (EFA)	15	38.5%
Item response theory (IRT)	2	5.1%
Model fit evaluation		
Yes	32	82.1%
No	7	17.9%

Note. The proportions are based on 39 study samples.

and IEA (https://ilsa-gateway.org/) databases of educational large-scale assessments. Four additional publications could be identified via personal contacts, email requests to authors, a snowball reference, and a university's thesis archive. Finally, to identify any further publications, we used the artificial intelligence-based tool "ASReview" (van de Schoot et al., 2021). Specifically, we submitted the outcomes of the PsycINFO and ERIC searches to the tool, screened 10% of the publications for their eligibility, and the tool ranked the remaining publications according to their similarity to the initially screened publications. The full documentation of the search can be accessed via https://bit.ly/3p517BF, and Fig. 1 shows the outcomes of these searches.

After removing duplicates, we screened the publications against the following criteria: (a) *Measures*: Primary studies included a quantitative measure of implicit theories/mindsets in intelligence; (b) *Coverage*: At least one type of mindset has been assessed with at least three items of the ITIS; (c) *Sample*: We included student samples enrolled in K-12 and university education (i) to align our meta-analysis with the extant literature providing evidence on the mindset-achievement relation and the effectiveness of mindset interventions in educational contexts (Costa & Faria, 2018; Sisk et al., 2018); and (ii) to cover the core levels of formal education (UNESCO Institute for Statistics, 2012). We excluded pre-K samples due to the issues of measuring mindsets by the ITIS in young children (Muradoglu et al., 2022); (d) *Statistical reporting*: Publications contain sufficient information to extract or derive the item-item correlations; (e) *Language*: Publications must be in English or, if not, authors have provided a summary of their study and the relevant information in English. In case of unpublished thesis and reports that did not undergo a peer review beyond institutional approval, we contacted the authors and asked for the raw data to retrieve the item-item correlations directly from the data. Overall, the literature search and screening procedures resulted in 29 reports of 32 primary studies which contained 39 independent samples (see Fig. 1). The full meta-analytic sample provided 487 correlation coefficients that were based on the item responses of 27,328 students.

3.2. Coding and effect sizes

As a next step, we extracted and coded study, sample, and measurement features to establish the contexts in which the ITIS had been used. Table 2 gives a detailed account of these features. Besides the features, we extracted and derived the item-item correlations as Pearson's *r*s from the primary studies. Specifically, some primary study authors have made available the raw data ($m = 7$), so that we could estimate the observed item-level correlation matrix directly (see Table 2). For $m = 11$ samples, the authors reported or provided us with the observed item-item correlation or covariance matrices, along with the descriptive statistics. Most study reports contained the results of factor analyses ($m = 21$), including factor correlations and item factor loadings. Applying (co-)variance rules, we derived the model-based item-item correlation matrices from the parameters of the exploratory factor analyses or the best-fitting confirmatory factor analysis models. In the case that authors have reported multiple factor models, we chose the parameters of the best-fitting model; in the case that authors had reported both CFA and EFA results, we chose the EFA results. The detailed procedure is described in the Supplementary Material S2. The agreement on the initial coding of the primary studies between two independent coders was 94%, and discrepancies were resolved subsequently.

3.3. Methodological approaches

Curran and Hussong (2009) noted that integrative data analysis represents an ideal approach to synthesizing item-level information and testing factor structures across multiple samples. This approach is based on premise that primary-study authors have made available their data sets—a premise that is largely not met in education (Logan et al., 2021). However, the recent advancements in item-level meta-analysis (Carpenter et al., 2016) and meta-analytic structural equation modelling (MASEM; Cheung, 2015a) allow meta-analysts to get close to the "ideal approach" by testing hypotheses on factor structures and moderators of model parameters based on item-item correlations. Drawing from these advancements, we (a) examined possible publication bias in the item-item correlations; (b) pooled the item-item correlation matrices via multivariate meta-analysis; (c) explored clusters of items via network analysis; (d) tested the factor structures of the two ITIS versions and their invariance across subgroups of primary studies; and (e) tested for moderator effects on the model parameters via one-stage MASEM.

3.3.1. Univariate meta-analyses

Before synthesizing correlation matrices across primary studies (RQ1), we evaluated possible publication bias in the single correlations. First, we pooled the single item-item correlations via univariate random-effects models with or without robust variance estimation (RVE) in the R packages "metafor" (Viechtbauer, 2010) and "robumeta" (Fisher et al., 2017). We chose RVE random-effects models as the primary models for the univariate meta-analyses to account for the hierarchical nesting of multiple, independent samples within primary studies. Second, we performed the funnel plot test by extending these models by the sample size as a moderator. Similarly, we performed Egger's Precision-Effect Test (PET; with the sampling standard error as moderator) and the Precision-Effect Estimate with Standard Errors (PEESE; with the sampling variance as moderator) in a third step. Fourth, we performed Begg's rank correlation test, followed by trim-and-fill analyses. To rule out file-drawer issues, we examined the *p*-curves of each correlation (Simonsohn et al., 2014). Finally, we identified influential correlations in the meta-analytic sample via Viechtbauer and Cheung's (2010) diagnostic procedure. Supplementary Material S3 shows the respective R code.

3.3.2. Multivariate meta-analyses

Given that each primary study contributed multiple item-item correlation coefficients nested in correlation matrices, the meta-analytic data exhibit effect size multiplicity (López-López et al., 2018). To address this multiplicity, we performed multiple approaches of multivariate meta-analysis to pool the correlation matrices across studies and quantify their heterogeneity. These

approaches included one-stage and two-stage meta-analytic structural equation modeling (OSMASEM and TSMASEM; Cheung, 2015a; Jak & Cheung, 2020) and meta-analytic aggregation of Gaussian networks (MAGNA; Epskamp et al., 2022).

Specifically, to address RQ1, we performed the first stage of TSMASEM, in which the correlation matrices R_i were synthesized across studies $i = 1, …, I$ to a pooled correlation matrix P via a multivariate random-effects model. Vectorizing these matrices to vectors r_i and ρ_R (Cheung, 2015a), this model is specified as $r_i = \rho_R + u_i + e_i$, with sampling errors e_i and study-specific deviations u_i from the pooled correlation matrix under normality assumptions, $e_i \sim MVN(0, V_i)$ and $u_i \sim MVN(0, T^2)$. V_i represents the sampling covariance matrix of study i, and T^2 the heterogeneity covariance matrix. Given the limited number of studies in our meta-analysis, we constrained the covariances among random effects in T^2 (i.e., the off-diagonals) to zero. The multivariate pooling approach accounts for the dependencies among correlation coefficients instead of assuming that they are independent—this results in a more accurate, pooled correlation matrix than separate meta-analysis of single correlations (Cheung, 2013).

To further address RQ2, we performed TSMASEM and MAGNA. We utilized the pooled correlation matrix from stage 1 in TSMASEM and submitted it to stage 2. Specifically, we estimated exploratory and confirmatory factor analysis models on the basis of the stage-1 correlation matrix and weights that corresponded to the inverse sampling variances and covariances via weighted least squares estimation (Cheung, 2015a). To supplement these correlation-based analyses, we further identified the relationships among mindset items via random-effects MAGNA, that is, analyses based on the partial correlations. MAGNA pools correlation matrices and estimates a multi-group Gaussian Graphical Model based on the corresponding, partial correlation matrices in one stage via maximum-likelihood estimation (Epskamp et al., 2022). Ultimately, MAGNA results in a network of items (nodes) and their partial correlations (exhibited as edges). Unlike factor analysis in which the relationships among two items are assumed to have a common cause (i.e., latent variables), the partial correlations used in MAGNA represent conditional dependencies among items after controlling for all other items (Epskamp et al., 2017). In this sense, the relations between two items represent direct dependencies accounting for the shared dependencies with other items in the network (Kan et al., 2019). Hence, MAGNA provides information about possible item connections from a different perspective than meta-analytic factor analysis and may even result in a better representation of the ITIS scale structure (e.g., Kan et al., 2019).

Addressing RQ3, we conducted subgroup analyses in the TSMASEM approach and tested the measurement invariance of the ITIS factor models. To study the specific moderator effects of sample, study, and measurement characteristics on the parameters in the measurement models, we also performed OSMASEM via maximum-likelihood estimation (Jak & Cheung, 2020). We performed TSMASEM and OSMASEM in the R package "metaSEM" (Cheung, 2015b) and MAGNA in "psychonetrics" (Epskamp, 2021).

3.3.3. Structural equation modelling

To evaluate the fit of the CFA models (RQ2), we referred to the traditional cut-offs of fit indices for an acceptable model fit (e.g., Hu & Bentler, 1999): an insignificant χ^2-value (indicating no major discrepancy between the observed and model-implied covariance matrices), the Comparative Fit Index (CFI) greater than or equal to 0.95, the Root Mean Square Error of Approximation (RMSEA) smaller than or equal to 0.06, and the Standardized Root Mean Squared Residual (SRMR) smaller than or equal to 0.08. However, as these recommendations have only been validated for a limited set of conditions, we did not consider them "golden rules" (see also Marsh et al., 2004). McNeish and Wolf (2021) developed an alternative, simulation-based procedure to adapt fit index cut-offs to the specific structural equation model and the context of the data. We report these dynamic cut-offs for the factor models relevant to RQ2. For subsequent invariance testing across subgroups of studies (RQ3), we estimated a series of multi-group CFA models with sequentially increasing constraints of model parameters. These models represented configural, metric, and structural invariance models and were compared via likelihood-ratio tests (Putnick & Bornstein, 2016).

3.4. Transparency and openness

To ensure the transparency and replicability of our research approaches and findings, we took three steps: (a) *Study*
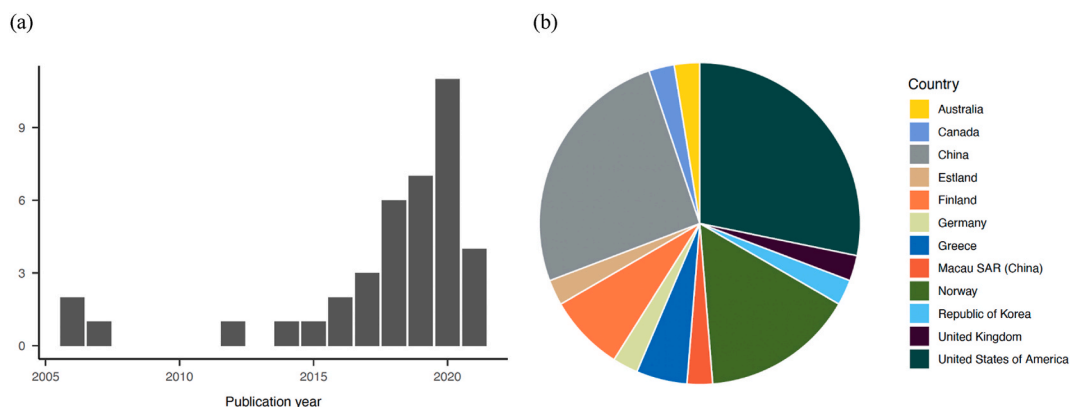
(a)

(b)

Fig. 2. Distribution of study samples per (a) publication year and (b) country.

preregistration—we pre-registered the present study within the Open Science Framework (OSF), including the study goals, analytic approaches, search strategies, and screening procedures of primary literature (https://bit.ly/3p517BF); (b) *Creating an open-access project*—we created an open-access project within the OSF to disclose the analytic code, output files, and the data set underlying our study (https://bit.ly/3DzfwKm); (c) *Providing supplementary material*—to ensure that readers can access the data, analytic code, and output files directly, we submitted the respective files as supplementary material attached to this journal article (see Supplementary Material S1–S7).

4. Results

4.1. Description of the meta-analytic sample, publication bias, and influential correlations

Table 2 and Fig. 2 detail the features of the meta-analytic sample. Most of the study reports were published in peer-reviewed journals or books and focused on validating the ITIS or describing the associations between mindset and other constructs ($m = 34$), yet hardly on the evaluation of interventions or changes over time ($m = 5$). Hence, the study designs were primarily cross-sectional ($m = 32$). The corresponding study samples contained students in secondary schools ($m = 24$) or universities ($m = 15$) who were, on average, 17.7 years old ($SD = 4.4$, $Mdn = 17.2$). The average proportion of boys in the study samples was 40.7% ($SD = 11.9$, $Mdn = 42.3$). Concerning the ITIS, the primary studies varied in the features of the scale: For instance, the number of response options ranged between 4 and 7, and most studies utilized the original six-point agreement scale (from "strongly disagree" to "strongly agree"). The number of items administered to the students ranged between 3 and 8 items. More than 60% of the study samples took the ITIS in a paper-and-pencil format ($m = 24$) rather than a computer-based (online) format ($m = 15$). Most studies included the ITIS in a way that both fixed and growth mindset could be assessed ($m = 24$); studies including only one mindset subscale assessed mainly fixed mindset ($m = 14$). Two-third of the samples have taken the ITIS with the original item formulations ($m = 26$). Concerning the measurement quality, reliability coefficients were by and large accessible ($m = 33$), and authors mainly chose to represent students' mindsets by latent variables ($m = 25$), testing measurement models with up to three factors. The primary approaches to representing mindsets were confirmatory and exploratory factor analyses. For about 82% of the study samples, model fit indices were evaluated to indicate the goodness of fit.

As an initial step before the pooling of correlation matrices, we evaluated the evidence on possible publication bias for the single

Table 3
Pooled item-item correlations matrix and heterogeneity indices for the six-item ITIS.

Item 2 (R)	Item 1 (R)	Item 2 (R)	Item 3 (R)	Item 4	Item 5
$\bar{r}$	.68				
95% $CI_{\bar{r}}$	[.64, .72]				
τ^2	.013				
95% CI_{τ^2}	[.006, .020]				
I^2	94.5%				
Item 3 (R)					
$\bar{r}$	.58	.61			
95% $CI_{\bar{r}}$	[.54, .63]	[.57, .65]			
τ^2	.016	.014			
95% CI_{τ^2}	[.007, .024]	[.007, .022]			
I^2	94.5%	95.4%			
Item 4					
$\bar{r}$	.40	.41	.35		
95% $CI_{\bar{r}}$	[.34, .47]	[.35, .47]	[.28, .42]		
τ^2	.016	.016	.021		
95% CI_{τ^2}	[.004, .028]	[.005, .027]	[.006, .036]		
I^2	91.6%	91.7%	93.4%		
Item 5					
$\bar{r}$	.40	.44	.39	.61	
95% $CI_{\bar{r}}$	[.33, .47]	[.37, .50]	[.32, .46]	[.56, .67]	
τ^2	.016	.014	.017	.010	
95% CI_{τ^2}	[.004, .029]	[.003, .025]	[.004, .030]	[.001, .018]	
I^2	91.9%	91.1%	92.3%	87.5%	
Item 6					
$\bar{r}$	.40	.40	.35	.58	.66
95% $CI_{\bar{r}}$	[.33, .47]	[.34, .47]	[.28, .42]	[.52, .64]	[.62, .71]
τ^2	.019	.018	.020	.014	.006
95% CI_{τ^2}	[.005, .033]	[.005, .031]	[.006, .034]	[.003, .025]	[.001, .012]
I^2	93.0%	92.7%	93.2%	91.0%	83.8%

Note. The pooled correlation matrix is based on 309 correlation coefficients derived from 39 independent samples in 32 primary studies ($N = 27328$). $\bar{r}$ = Average correlation under the random-effects model, 95% $CI_{\bar{r}}$ = 95% confidence interval of $\bar{r}$, τ^2 = Between-sample heterogeneity with confidence interval 95% CI_{τ^2}, I^2 = Heterogeneity index. (R) = Items represented fixed mindsets and were treated as reverse-coded.

item-item correlations. The respective models and statistics are detailed in the Supplementary Material S3. Overall, these analyses showed that: (a) None of the extracted correlations were influential; (b) All of the *p*-values had evidential value; (c) Some publication bias was evident—the PET and PEESE procedures indicated possible publication bias for the correlations r_{13}, r_{17}, r_{23}, r_{27}, r_{37}, r_{58}, and r_{68}. However, all other tests did not flag these correlations. We therefore argue that the degree of publication bias was small.

4.2. Multivariate pooling of correlation matrices (RQ1)

To address RQ1, we pooled the item-item correlation matrices via stage-1 TSSEM under the assumptions of fixed or random effects. Notably, the six-item ITIS data were comprised of 6 × 6 correlation matrices, while the eight-item ITIS data were comprised of 8× 8 correlation matrices. For both data sets, the pooling via a fixed-effects model resulted in poor fit (six-item ITIS: χ^2 [294] = 4940.1, $p <$ .001, CFI = 0.906, RMSEA = 0.150, 95% CI [0.147, 0.154], SRMR = 0.140; eight-item ITIS: χ^2 [459] = 5971.41, $p <$ .001, CFI = 0.914, RMSEA = 0.131, 95% CI [0.128, 0.134], SRMR = 0.140). This observation indicated that the assumption of fixed effects does not hold. Indeed, the pooling of correlation matrices under a random-effects model indicated substantial heterogeneity in all correlations (six-item ITIS: $I^2 = 83.8$–95.4%; eight-item ITIS: $I^2 = 76.5$–95.2%). Hence, we accepted the pooled item-item correlations from this model.

For the six-item ITIS, the pooled correlations ($\bar{r}$) among the fixed-mindset items ranged between 0.58 and 0.66, and between 0.58 and 0.68 for the growth-mindset items (see Table 3). The correlations between fixed- and growth mindset items were also positive yet

Table 4
Pooled item-item correlations matrix and heterogeneity indices for the eight-item ITIS.

	Item 1 (R)	Item 2 (R)	Item 3 (R)	Item 4	Item 5	Item 6	Item 7 (R)
Item 2 (R)							
$\bar{r}$	.68						
95% $CI_{\bar{r}}$	[.64, .72]						
τ^2	.013						
95% CI_{τ^2}	[.006, .019]						
I^2	94.3%						
Item 3 (R)							
$\bar{r}$	.58	.61					
95% $CI_{\bar{r}}$	[.54, .63]	[.57, .65]					
τ^2	.015	.014					
95% CI_{τ^2}	[.007, .023]	[.007, .021]					
I^2	94.2%	95.2%					
Item 4							
$\bar{r}$	.40	.41	.35				
95% $CI_{\bar{r}}$	[.34, .46]	[.35, .47]	[.28, .42]				
τ^2	.015	.014	.019				
95% CI_{τ^2}	[.004, .025]	[.004, .024]	[.006, .033]				
I^2	90.8%	90.9%	92.9%				
Item 5							
$\bar{r}$	.40	.43	.39	.61			
95% $CI_{\bar{r}}$	[.33, .47]	[.37, .49]	[.32, .46]	[.56, .67]			
τ^2	.015	.013	.016	.009			
95% CI_{τ^2}	[.003, .026]	[.003, .023]	[.004, .027]	[.001, .016]			
I^2	91.1%	90.2%	91.6%	86.0%			
Item 6							
$\bar{r}$	.40	.40	.35	.58	.66		
95% $CI_{\bar{r}}$	[.33, .47]	[.33, .47]	[.28, .42]	[.52, .64]	[.62, .71]		
τ^2	.018	.017	.018	.013	.005		
95% CI_{τ^2}	[.005, .032]	[.005, .029]	[.005, .031]	[.003, .023]	[.001, .010]		
I^2	92.7%	92.2%	92.6%	90.3%	81.9%		
Item 7 (R)							
$\bar{r}$	.62	.65	.57	.44	.45	.45	
95% $CI_{\bar{r}}$	[.56, .68]	[.59, .70]	[.50, .64]	[.34, .54]	[.35, .55]	[.36, .55]	
τ^2	.014	.011	.019	.025	.023	.022	
95% CI_{τ^2}	[.004, .025]	[.003, .019]	[.005, .033]	[.003, .048]	[.001, .044]	[.002, .042]	
I^2	91.4%	89.0%	93.1%	94.4%	93.9%	93.7%	
Item 8							
$\bar{r}$	.40	.43	.41	.57	.64	.65	.44
95% $CI_{\bar{r}}$	[.33, .47]	[.36, .50]	[.33, .50]	[.52, .63]	[.59, .68]	[.60, .70]	[.35, .54]
τ^2	.014	.013	.023	.008	.005	.005	.024
95% CI_{τ^2}	[.002, .026]	[.002, .024]	[.004, .042]	[.001, .015]	[.000, .010]	[.000, .011]	[.003, .045]
I^2	90.4%	89.9%	93.9%	84.8%	76.5%	78.5%	94.2%

Note. The pooled correlation matrix is based on 487 correlation coefficients derived from 39 independent samples in 32 primary studies ($N = 27328$). $\bar{r}$ = Average correlation under the random-effects model, 95% $CI_{\bar{r}}$ = 95% confidence interval of $\bar{r}$, τ^2 = Between-sample heterogeneity with confidence interval 95% CI_{τ^2}, I^2 = Heterogeneity index. (R) = Items represented fixed mindsets and were treated as reverse-coded.

lower, $\bar{r} = 0.35$-0.44 (see Table 4). Similar ranges occurred for the eight-item ITIS (fixed-mindset items: $\bar{r} = 0.57$-0.68; growth-mindset items: $\bar{r} = 0.57$-0.66; fixed- and growth mindset items: $\bar{r} = 0.35$-0.45). We utilized these pooled correlation matrices for the subsequent factor analyses of the full meta-analytic sample and generated group-specific correlation matrices for the subgroup analyses, following the same procedure. Notably, in both MAGNA and OSMASEM, the pooling of correlation matrices and the estimation of the analytic models are performed at once.

4.3. Meta-analytic network models

As a further step towards addressing RQ2, we examined the connections between the ITIS items via meta-analytic Gaussian network aggregation (MAGNA). For both the six- and eight-item ITIS, random-effects MAGNA models represented the data well and were used to make visible the partial item-item correlations (six-item ITIS: CFI = 1.00, RMSEA = 0.00, 90% CI [0.00, 0.00]; eight-item ITIS: CFI = 1.00, RMSEA = 0.00, 90% CI [0.00, 0.11]). Supplementary Material S6 details these models.

Six-item ITIS. The partial correlations ω_{ij} between the fixed mindset items i and j ranged from 0.25 to 0.43 and from 0.26 to 0.41 between the growth mindset items; partial correlations between fixed and growth mindset items were substantially lower, $\omega = 0.04$-0.09 ($ps \leq .002$). Moreover, the resultant values of the expected influence within the network ranged between 0.74 (item 3) and 0.97 (item 5). Fig. 3a depicts the overall network and indicates that two clusters of items existed, one comprising the fixed-mindset items (items 1–3) and one comprising the growth mindset items (items 4–6). The connections between these clusters were all positive, yet weak.

Eight-item ITIS. The partial correlations were similar to those for the six-item ITIS (between fixed mindset items: $\omega = 0.19$-0.34, between growth mindset items: $\omega = 0.14$-0.32, between fixed and growth mindset items: $\omega = 0.02$-0.19; $ps \leq .085$). The expected influence values ranged between 0.74 (item 4) and 1.02 (item 5) and identified items 2, 5, and 7 as the most important network nodes. Similar to the 6-item ITIS network, two item clusters corresponded directly to the two mindset subscales (see Fig. 3b). These clusters were positively yet weakly connected. Notably, the partial correlation between items 4 and 8 was considerably larger than the other between-cluster partial correlations, $\omega_{48} = 0.19$. Consequently, the subsequent factor analyses may contain cross-loadings or lower within-scale factor loadings involving items 4 and 8.

Overall, the network analyses suggested the existence of two connected but distinct item clusters that corresponded to fixed and growth mindset. The between-cluster connections were stronger for the eight-item ITIS than for the six-item ITIS.

4.4. Meta-analytic structural equation modeling (MASEM)

4.4.1. Exploratory factor analyses (RQ2)

Addressing RQ2, we first performed exploratory factor analyses on the pooled correlation matrices. For both the six- and eight-item ITIS, the Kaiser-Mayer-Olkin test provided mean sampling adequacies above 0.80, and Bartlett's sphericity test resulted in significant chi-square statistics (see Table 5). Hence, the correlation matrices supported that the items were correlated, and there was sufficient common variance (Kaiser & Rice, 1974). The Empirical Kaiser Guttman criteria for the eigenvalues indicated two rather than one factor (Braeken & van Assen, 2017; see Supplementary Material S4 and S5). The factor analysis with Oblimin rotation resulted in a factor correlation of $\rho = 0.63$ for the six-item ITIS, and $\rho = 0.64$ for the eight-item ITIS, respectively, and exhibited good fit to the data (see Table 5). The factor structure underlying the six-item ITIS was close to a simple structure, with items 1–3 loading on one factor, items 4–6 loading on another factor, and small cross-loadings ($\lambda = -0.02$–0.05). Similarly, items 1–3 and 7 represented one factor, while items 4–6 and 8 represented another factor. However, cross-loadings ranged up to $\lambda = .11$ (item 7).

Taken together, the exploratory factor analyses suggested two correlated factors describing the structure of the six- and eight-item
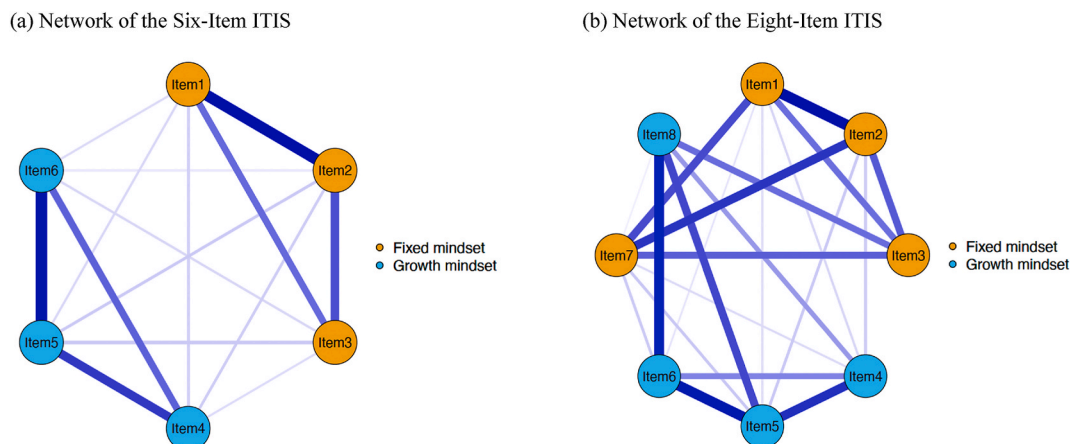

Fig. 3. Random-Effects Meta-Analytic Gaussian Network Aggregation
Note. ITIS = Implicit Theories of Intelligence Scale. Items representing fixed mindsets were treated as reverse-coded.

Table 5

Results from the meta-analytic exploratory factor analyses of the six- and eight-item ITIS.

Item	Six-Item ITIS			Eight-Item ITIS		
	Factor loading		Communality	Factor loading		Communality
	Factor 1	Factor 2	h^2	Factor 1	Factor 2	h^2
Item 1 (R)	**.81**	.00	.65	**.83**	−.03	.65
Item 2 (R)	**.85**	.00	.71	**.86**	−.03	.71
Item 3 (R)	**.71**	.02	.52	**.72**	.01	.53
Item 4	.05	**.70**	.54	.06	**.69**	.54
Item 5	−.01	**.85**	.70	−.01	**.82**	.67
Item 6	−.02	**.80**	.63	−.04	**.84**	.66
Item 7 (R)	–	–	–	**.71**	.11	.61
Item 8	–	–	–	.03	**.77**	.62
Model results						
Factor correlation	.63			.64		
Overall MSA	.83			.90		
Bartlett's test of sphericity χ^2 (df)	72657.2 (15)*			114349.2 (28)*		
RMSR	.01			.01		
RMSEA	.031			.039		
90% CI_{RMSEA}	[.026, .036]			[.036, .042]		

Note. N = 27328. The factor analyses were based on Oblimin rotation and maximum-likelihood estimation. (R) = Items represented fixed mindsets and were treated as reverse-coded. Factor loadings above 0.30 are in bold. MSA = Measure of sampling adequacy (Kaiser & Rice, 1974), RMSR = Root Mean Square of the Residuals, RMSEA = Root Mean Square Error of Approximation, *df* = degrees of freedom, ITIS = Implicit Theories of Intelligence Scale. *$p < .001$.

ITIS. Some items in the eight-item ITIS exhibited cross-loadings and may cause model fit deterioration (Li et al., 2020).

4.4.2. Confirmatory Factor Analyses (RQ2)

To further substantiate the evidence supporting the two-factor model, we specified a single-factor and a two-factor model within the CFA framework and compared the respective model fit indices. Table 6 details the fit indices and the results of the chi-square difference tests for the full meta-analytic samples. Despite the observation that the two-factor models did not only meet the commonly used model fit index cutoffs presented by Hu and Bentler (1999), they also met the criteria of the dynamic model fit index cutoffs, thus pointing to their good representation of the meta-analytic data (i.e., dynamic model fit index cutoffs: six-item ITIS: CFI = 0.979, RMSEA = 0.102, SRMR = 0.036; eight-item ITIS: CFI = 0.972, RMSEA = 0.090, SRMR = 0.044). For both the six- and eight-item ITIS, a two-factor model with two correlated latent variables representing fixed and growth mindsets (see Fig. 4) exhibited very good model fit and outperformed the corresponding single-factor models. Treating the fixed-mindset items as reverse-coded, the resultant factor correlations were $\rho = 0.633$ (95% CI [0.589, 0.677]) for the six-item ITIS, and, respectively, $\rho = 0.651$ (95% CI [0.616, 0.686]) for the eight-item ITIS. Fig. 3 depicts the underlying, meta-analytic CFA models and further shows the high factor loadings for both mindsets, $\lambda_F = 0.72$-0.84 and $\lambda_G = 0.74$-0.83. Ultimately, these two latent variables were highly reliable in both versions of the ITIS (six-item ITIS: McDonald's $\omega_F = 0.84$, $\omega_G = 0.83$; eight-item ITIS: $\omega_F = 0.87$; $\omega_G = 0.87$). Overall, the meta-analytic CFA supported the preference of the two-factor model representing students' implicit theories of intelligence.

4.4.3. Multi-group Confirmatory Factor Analyses (RQ3)

Consistently across subgroups of study samples (i.e., type of response scale, scale modifications, assessment mode, educational level and origin of the sample), the two-factor model fitted the data well and outperformed the single-factor model in model fit (see Appendix Table A1). The respective factor correlations ranged between $\rho = 0.42$ (subgroup modified scales) and 0.77 (subgroup

Table 6

Model fit indices and comparisons resulting from confirmatory factor analyses.

Model	χ^2			RMSEA		SRMR	CFI	AIC	BIC	$\Delta\chi^2$		
	Value	df	p	Value	95% CI					Value	Δdf	p
Six-Item ITIS												
Full sample (N = 27328, m = 39, l = 309)												
Single-factor model	208.5	9	<.001	.029	[.025, .032]	.123	.955	190.5	116.5	–	–	–
Two-factor model	1.7	8	.99	.000	[.000, .000]	.011	1.000	−14.3	−80.0	206.8	1	<.001
Eight-Item ITIS												
Full sample (N = 27328, m = 39, l = 487)												
Single-factor model	318.5	20	<.001	.023	[.021, .026]	.116	.959	278.5	114.2	–	–	–
Two-factor model	8.7	19	.98	.000	[.000, .000]	.022	1.000	−29.3	−185.4	309.8	1	<.001

Note. RMSEA = Root Mean Square Error of Approximation, *df* = degrees of freedom, 95% CI = 95% confidence interval, SRMR = Standardized Root Mean Squared Residual, CFI = Comparative Fit Index, AIC = Akaike's Information Criterion, BIC = Bayesian Information Criterion, N = Overall size of the study samples, *m* = Number of independent samples, *l* = Number of correlation coefficients, ITIS = Implicit Theories of Intelligence Scale.

(a) Six-Item ITIS

(b) Eight-Item ITIS
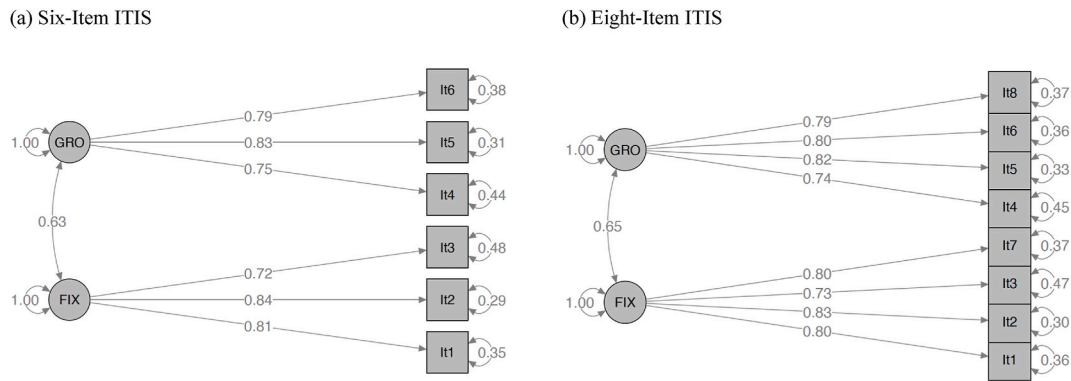
Fig. 4. Meta-Analytic Confirmatory Factor Analysis via Two-Stage MASEM
Note. ITIS = Implicit Theories of Intelligence Scale, GRO = Growth mindset, FIX = Fixed mindset, It = Item. Items representing fixed mindsets were treated as reverse-coded.

Table 7
Results of the measurement invariance testing across subgroups.

Grouping variable	Model	Deviance ($-2\log\mathscr{L}$)	Npar	AIC	BIC	χ^2_{LRT} Value	Δdf	p
Six-Item ITIS								
Educational level (*1 = School, 0 = University*)	Configural	66.6	26	118.6	332.2	–	–	–
	Metric	396.9	20	436.9	601.2	330.3	6	<.001
	Structural	632.5	19	670.5	826.6	565.9	7	<.001
Type of response scale (*1 = Six-point, 0 = Others*)	Configural	4.2	26	56.2	269.8	–	–	–
	Metric	6.8	20	46.8	211.1	2.6	6	.85
	Structural	6.8	19	44.8	200.9	2.6	7	.92
Assessment mode (*1 = Online, 0 = Paper-pencil*)	Configural	10.0	26	62.0	275.6	–	–	–
	Metric	14.4	20	54.4	218.7	4.4	6	.62
	Structural	42.1	19	80.1	236.2	32.1	7	<.001
Scale modifications (*1 = Modified, 0 = Original scale*)	Configural	30.8	26	82.8	296.4	–	–	–
	Metric	115.8	20	155.8	320.1	85.0	6	<.001
	Structural	264.2	19	302.2	458.3	233.4	7	<.001
East-Asian samples (*1 = East-Asian, 0 = Others*)	Configural	104.4	26	156.4	370.0	–	–	–
	Metric	151.4	20	191.4	355.8	47.0	6	<.001
	Structural	209.4	19	247.4	403.5	105.0	7	<.001
Eight-Item ITIS								
Educational level (*1 = School, 0 = University*)	Configural	206.6	34	274.6	554.0	–	–	–
	Metric	674.3	26	726.3	939.9	467.6	8	<.001
	Structural	1008.7	25	1058.7	1264.1	802.1	9	<.001
Type of response scale (*1 = Six-point, 0 = Others*)	Configural	15.6	34	83.6	362.9	–	–	–
	Metric	19.4	26	71.4	285.0	3.9	8	.87
	Structural	19.5	25	69.5	274.9	3.9	9	.92
Assessment mode (*1 = Online, 0 = Paper-pencil*)	Configural	23.4	34	91.4	370.8	–	–	–
	Metric	34.0	26	86.0	299.5	10.6	8	.23
	Structural	59.3	25	109.3	314.7	35.9	9	<.001
Scale modifications (*1 = Modified, 0 = Original scale*)	Configural	164.9	34	232.9	512.2	–	–	–
	Metric	312.0	26	364.0	577.6	147.1	8	<.001
	Structural	485.5	25	535.5	740.9	320.7	9	<.001
East-Asian samples (*1 = East-Asian, 0 = Others*)	Configural	171.6	34	239.6	518.9	–	–	–
	Metric	228.8	26	280.8	494.4	57.2	8	<.001
	Structural	311.6	25	361.6	567.0	140.1	9	<.001

Note. $\log\mathscr{L}$ = Log-likelihood value, Npar = Number of parameters, AIC = Akaike's Information Criterion, BIC = Bayesian Information Criterion, χ^2_{LRT} = Chi-square value with Δdf degrees of freedom based on the likelihood-ratio test (LRT). Model comparisons are conducted against the configural model.

university samples) for six-item ITIS, and between $\rho = .44$ (subgroup modified scales) and 0.80 (subgroup university samples) for the eight-item ITIS (see Appendix Table A2). The result that the two-factor model was superior within all the a-priori defined subgroups indicated a high level of robustness of this finding. Similar to the full meta-analytic sample, scale reliabilities were high (six-item ITIS: $\omega_F = .78$-.90 and $\omega_G = 0.79$-0.90; eight-item ITIS: $\omega_F = 0.81$-0.93 and $\omega_G = 0.83$-0.92). Overall, the two-factor model provided reliable factors and qualified as a baseline model for further invariance and moderation tests.

To further examine possible moderator effects, we tested the two-factor model for its configural, metric, and structural invariance via multi-group CFA. As it is common in invariance testing, we compared the fit of the two latter invariance models to that of the configural invariance model (Putnick & Bornstein, 2016). Table 7 contains the results of these comparisons (see also Supplementary Material S4 and S5). For both ITIS versions, only configural invariance was achieved across educational levels, scale modifications, and East-Asian original samples. We have obtained evidence supporting metric invariance across assessment modes, and even structural invariance across types of response scales. These results point to the moderating roles educational levels, scale modifications, and East-Asian origin samples play for factor loadings and the factor correlation; besides, assessment mode exhibited moderator effects on the factor correlation. Only the type of response scale did not moderate any model parameter.

4.4.4. Moderator effects via one-stage MASEM (RQ3)

The meta-analytic and multi-group CFA provided information about the moderator effects of the sample, study, and measurement features on the factor correlation and the set of factor loadings. However, they neither identified which loadings might be specifically affected nor the direction of these effects. We therefore performed OSMASEM to supplement this information. These analyses revealed the following (see Table 8):

- *Educational level*: For both the six- and eight-item ITIS, negative moderator effects on factor loadings existed for all items, indicating that school samples resulted in lower loadings. In addition, the factor correlations were moderated, with higher coefficients for university samples.
- *East-Asian samples*: Samples originating from East-Asian countries showed lower factor correlations. There was no consistent evidence on the moderation of factor loadings for both ITIS versions, except for a positive effect on the loading of item 3 in the eight-item ITIS.
- *Type of response scale*: This feature did not show any moderator effects.

Table 8
Moderator effects on the parameters in the two-factor measurement model.

Moderator	Moderation effects on factor loadings λ_{im}				Moderation effects on the factor correlation ρ_m			
	χ^2_{LRT}	Δdf	p	Effects B_i	χ^2_{LRT}	Δdf	p	Effect B
Six-Item ITIS								
Educational level (1 = School, 0 = University)	116.1	6	<.001	Items 1–6: $Bs = -.17$ to $-.06$, $ps \leq .014$	167.0	7	<.001	$B = -0.25$, $SE = 0.03$, $p < .001$
Type of response scale (1 = Six-point, 0 = Others)	2.1	6	.91	Items 1–6: $Bs = -.03 - .02$, $ps \geq .28$	2.7	7	.91	$B = 0.04$, $SE = 0.05$, $p = .42$
Assessment mode (1 = Online, 0 = Paper-pencil)	10.7	6	.10	Item 4: $B_4 = 0.09$, $SE = 0.05$, $p = .07$; Items 1–3 & 5: $Bs = .01$–.06, $ps \geq .23$	24.9	7	<.001	$B = 0.16$, $SE = 0.04$, $p < .001$
Scale modifications (1 = Modified, 0 = Original scale)	52.7	6	<.001	Items 3–5: $Bs = -.16$ to $-.11$, $ps \leq .016$; Items 1–2, & 6: $Bs = -.07$ to $-.05$, $ps \geq .16$	66.5	7	<.001	$B = -0.18$, $SE = 0.04$, $p < .001$
East-Asian samples (1 = East-Asian, 0 = Others)	7.0	6	.33	Items 1–6: $Bs = -.03 - .02$, $ps \geq .09$	11.6	7	.11	$B = -0.13$, $SE = 0.06$, $p = .026$
Eight-Item ITIS								
Educational level (1 = School, 0 = University)	162.9	8	<.001	Items 1–8: $Bs = -.18$ to $-.06$, $ps \leq .008$	286.0	9	<.001	$B = -0.32$, $SE = 0.03$, $p < .001$
Type of response scale (1 = Six-point, 0 = Others)	4.0	8	.86	Items 1–8: $Bs = -.01 - .05$, $ps \geq .13$	4.0	9	.91	$B = 0.01$, $SE = 0.04$, $p = .86$
Assessment mode (1 = Online, 0 = Paper-pencil)	17.2	8	.03	Item 4: $B_4 = 0.08$, $SE = 0.03$, $p = .02$; Items 1–3 & 5–8: $Bs = -.01 - .05$, $ps \geq .12$	37.5	9	<.001	$B = 0.15$, $SE = 0.03$, $p < .001$
Scale modifications (1 = Modified, 0 = Original scale)	96.6	8	<.001	Items 3–5 & 7–8: $Bs = -.21$ to $-.07$, $ps \leq .055$; Items 1–2, & 6: $Bs = -.05$ to $-.04$, $ps \geq .24$	110.3	9	<.001	$B = -0.15$, $SE = 0.04$, $p < .001$
East-Asian samples (1 = East-Asian, 0 = Others)	12.1	8	.15	Item 3: $B_3 = 0.08$, $SE = 0.04$, $p = .04$; Items 1–2 & 4–8: $Bs = -.10 - .03$, $ps \geq .067$	20.5	9	.02	$B = -0.15$, $SE = 0.05$, $p < .01$

Note. The moderator effects were estimated via one-stage MASEM. χ^2_{LRT} = Chi-square value with Δdf degrees of freedom based on the likelihood-ratio test (LRT). The LRT was based on the comparison between the two-factor CFA models with or without moderators of the factor loadings or factor loadings and the factor correlation. i = Item number, m = Number of the study sample, ITIS = Implicit Theories of Intelligence Scale.

- *Assessment mode*: The factor correlation was moderated positively for both the six- and eight-item ITIS, with larger coefficients resulting from computer-based (online) assessments. In the six-item ITIS, we observed a tendency toward a positive moderation of the factor loading of item 4; this tendency was evident and significant for the eight-item ITIS. No further moderator effects could be found.
- *Scale modification*: Consistently across the two ITIS versions, the factor loadings of items 1, 2, and 6 were not moderated; yet those of items 3–5, with negative effects indicating lower loadings in primary studies with modified scales. Items 7 and 8 followed the same moderator pattern. Moreover, significantly higher factor correlations could be observed in studies administering the original ITIS.

By and large, OSMASEM supported the previous invariance results and highlighted some items that may function differently across the selected features.

5. Discussion

5.1. Structure and heterogeneity of the ITIS

As we set out to examine key properties of the two ITIS versions as the most popular assessments of mindsets in education, we discovered (a) overall positive item-item correlations with substantial heterogeneity (RQ1); (b) the preference of measurement models with two moderately to highly correlated factors representing the incremental and entity theories of intelligence (RQ2); and (c) the moderation of model parameters (i.e., factor loadings and correlation) by the educational level and origin of the sample, assessment mode, and scale modifications (RQ3). Table 9 summarizes these findings in greater detail.

As expected, the positive and at least moderate item-item correlations suggested homogeneity within each mindset subscale and point to the existence of some underlying construct (Borsboom et al., 2003). Not surprisingly, the correlations among items measuring different mindsets were weaker—this finding remained even after reversing one set of mindset items to align with the other and was robust against all conditions and forms of grouping in our meta-analytic sample. If the "between-mindsets" correlations are indeed weaker than the "within-mindsets" correlations, then the existence of two rather than one latent variable representing the underlying constructs is indicated. In fact, the network analyses pointed into the same direction for both the six- and eight-item version of the ITIS, that is, two clusters of items occurred, each of which represented one mindset, and positive item-item connections within these clusters. We notice that the meta-analytic aggregation of networks via MAGNA provided a useful tool to explore the item-item connections based on partial correlations and, at the same time, quantify heterogeneity (Epskamp et al., 2022).

All item-item correlations in both ITIS versions exhibited between-sample heterogeneity. This heterogeneity may be interpreted in two ways: First, the item-item correlation could not be fully replicated across samples (Hedges & Schauer, 2019). However, the heterogeneity in item-item correlations does not necessarily imply that the core finding of our study, that is, the existence of two correlated mindset factors, is also not replicable. Second, the between-sample variation may be linked to sample, study, or measurement characteristics—information about these links may be especially useful for informing the design of future studies and benchmarking their results (Pigott & Polanin, 2020). The finding that educational level, the origin of the student samples, and measurement characteristics showed moderation effects was in line with existing meta-analyses of the mindset-achievement relation (Costa & Faria, 2018; Sisk et al., 2018). Nevertheless, our list of possible moderators was not exhaustive. For instance, students' gender may explain parts of the remaining heterogeneity. In fact, some empirical studies suggested significant gender differences in mindsets (e.g., Macnamara & Rupani, 2017; OECD, 2021). However, the gender composition of the student samples reviewed by Costa and Faria (2018) did not moderate the mindset-achievement relation, and some evidence points to the measurement invariance of mindset scales across gender (Bostwick et al., 2017) and small, if not insignificant gender differences in mindset scores (e.g., OECD, 2021; Sigmundsson et al., 2021). Moreover, students' cognitive abilities may also moderate the parameters in the mindset measurement model. In a large-scale study of the Rosenberg Self-Esteem Scale, Gnambs and Schroeders (2017) showed that several measurement properties

Table 9
Overview of the key findings.

Research Question (RQ)	Key Findings
RQ1. To what extent are the ITIS items correlated, and how do these correlations vary across study samples? (*Pooled correlations and heterogeneity*)	•Positive correlations among fixed-mindset items, $\bar{r}$ = .57–.68 •Positive correlations among growth-mindset items, $\bar{r}$ = .57–.68 •Positive but lower correlations between fixed- and growth-mindset items, $\bar{r}$ = .35–.45 •Substantial heterogeneity in item-item correlations between samples, I^2 = 76.5–95.4%
RQ2. To what extent can the single- and two-factor models represent the structure of the ITIS, and what characterizes the best-fitting model? (*Factor structures*)	•Preference of the two-factor model over the single-factor model for the full data sets •Preference of the two-factor model over the single-factor model consistently across subgroups of primary studies •Factor correlations between ρ = .42–.80 •High scale reliabilities, ω = .78–.93
RQ3. Which study, sample, and measurement characteristics moderate the model parameters in the final ITIS factor model? (*Moderation of model parameters*)	•No moderation by the types of response scales •Moderation of factor loadings and the factor correlation by the educational level and origin of the sample, assessment mode, and scale modifications

(i.e., dimensionality and wording effects) were associated with students' cognitive abilities, possibly via response styles. Given that the ITIS captures mindsets via similar self-reports, cognitive abilities may explain heterogeneity.

Until now, most meta-analyses describing the relations between implicit theories and other constructs have quantified and explained the heterogeneity in scale scores, assuming that the measurement models and item-item correlations are homogeneous across study samples (e.g., Costa & Faria, 2018; Sisk et al., 2018). Our meta-analysis, however, showcases that heterogeneity is, in fact, located at the level of item-item correlations which form the basis for measurement models in meta-analyses. This is where the potential of correlation-based MASEM at the level of items lies (Cheung, 2015a)—in quantifying and possibly explaining heterogeneity at the level of items, the key sources of information for the structure of mindset assessments.

Both versions of the ITIS exhibited a two-factor structure in the subsequent factor analyses for the full sample of primary studies and sub-samples. Finding that this structure, by and large, held within and across sub-groups of studies suggests its robustness and applicability in various contexts (Higgins et al., 2019). From a substantive perspective, this finding implies that students can hold both incremental and entity theories simultaneously. In their review of the conceptual and assessment issues of implicit theories, Lüftenegger and Chen (2017) argued that it may indeed be possible to observe both mindsets in achievement situations or in interventions. Recently, more and more evidence has accumulated backing a "mixed" mindset (e.g., Glerum et al., 2020; Li & Bates, 2020; Lou et al., 2021). These findings could impact how researchers conceptualize implicit theories, that is, no longer as mutually exclusive beliefs but as two beliefs that are connected.

Albeit the meta-analytic evidence suggests two correlated factors, it still needs to be further examined if they indeed represent the two implicit theory constructs. We argue that additional evidence is needed to back this distinction and strengthen the validity argument (Pellegrino et al., 2016). Examining the relations to other constructs, such as educational achievement and intelligence, across the two mindsets to avoid "jingle-jangle fallacies" (Gonzalez et al., 2021). In this way, evidence on the link between mindsets and their reference construct (i.e., intelligence) could support the crafting of a validity argument. A study by Macnamara and Rupani (2017) showed that such a link may exist. Such evidence is especially relevant because the multidimensionality of the ITIS could also be caused by the mixture of item wordings that are positive (e.g., incremental) or negative (e.g., entity) relative to the construct (e.g., Gnambs & Schroeders, 2017; Steinmann et al., 2021). If this applies, then the two correlated factors merely represent methodological artefacts that may be construct-irrelevant. Irrespective of the source of multidimensionality, the core finding that the ITIS entertains two correlated factors could impact the way in which researchers use this scale and draw inferences on students' mindsets.

5.2. Implications for using the implicit theories of Intelligence Scale

Our findings have several implications for the use of the ITIS: First, the existence of two correlated mindset factors even after reversing one item set challenges the practice of reverse coding and its assumption. Reverse-coding items does not necessarily align the construct meaning and lead to the unidimensionality of the ITIS (see also Lüftenegger & Chen, 2017). This applies to other constructs as well (Steinmann et al., 2021) and is a well-known issue in educational measurement (Kam, 2016). Hence, if educational researchers assess only one type of mindset, by multiple items or a single item, they should not draw inferences on the other type. We caution against the use of single-item measures, given their limited reliability and the possible response bias (Rammstedt et al., 2021). If researchers wish to capture implicit theories more broadly, they should assess both mindsets and draw inferences on these two constructs rather than a unified construct. This, however, requires measurement models that represent the two mindsets within the ITIS. In this situation, reporting a mindset profile is a suitable alternative to categorizing students into fixed vs. growth mindsets (Glerum et al., 2020; Lüftenegger & Chen, 2017).

Second, the factor and network analyses suggested that adding items 7 and 8 to the six-item ITIS improves sub-scale reliabilities, yet all other psychometric properties were similar. Both versions exhibited non-invariance across educational levels so that a preference of one over the other for specific school or university samples did not become clear. The six- and eight-item ITIS seem applicable to the educational levels we have studied in our meta-analysis, yet not fully comparable across these levels.

Third, we encourage researchers who wish to use the ITIS in their studies to consider the context in which the ITIS is administered. Specifically, we showed that the context of the student sample explains heterogeneity in the meta-analytic data and that most parameters in the measurement models were not invariant across sub-groups. The functioning of the ITIS seems sensitive to the sample context and requires measurement invariance testing if group comparisons of mindsets are of interest. Moreover, the observations that measurement characteristics explained heterogeneity suggests once more to avoid assuming that mindset scores are fully comparable across assessment modes or scale modifications. From our perspective, being consistent with the mindset assessment in studies with multiple samples or conditions, testing for possible non-invariance as part of group comparisons, and reporting transparently the context of the study and sample are key to further mindset research.

5.3. Limitations and future directions

Our study has several limitations: First, our meta-analysis focused on the assessment of mindsets in education via the Implicit Theories of Intelligence Scale, originally developed by Dweck (2000), yet neither included other scales nor mindset assessments focusing on mathematics, reading, personality, or other key educational domains. We therefore encourage meta-analysts to initiate and extend item-level meta-analyses—this could shed light on the possible differences and similarities of different mindset assessments and their functioning across domains. Second, some item-item correlations had to be retrieved from the parameters of factor models. Although we chose the best-fitting model or EFA over CFA to approximate the observed correlation matrix, our meta-analytic sample comprised model-implied and observed correlation matrices. The extent to which this mixture may create bias toward a specific factor

model in MASEM is to be clarified. Third, item-level meta-analyses require the transparent and sufficient reporting of psychometric properties, item-item correlations, or raw data in the primary studies. This requirement could create selection bias in the meta-analytic sample, preferring studies meeting the reporting requirements. Together with Plucker and Makel (2021), we therefore encourage educational researchers to adopt a culture of open science and transparency to circumvent such selection bias in subsequent meta-analyses.

6. Conclusion

Our item-level meta-analysis revealed that the most commonly used assessment of students' implicit theories of intelligence, the ITIS, measures two moderately correlated mindset factors, one corresponding to incremental theories and one to entity theories of intelligence. This finding held even after reverse-coding the item set of one of these theories and applied to all subgroups of primary studies, that is, across educational levels, origins of the student sample, assessment modes, and scale modifications. From the perspective of educational measurement, our study challenges current assessment practices of mindsets, especially the reverse-coding of fixed- or growth-mindset items under the assumption that the meanings of the items align after the reversing. However, it still needs to be clarified whether the two mindset factors also show differential relations to educational outcomes, such as academic achievement, and whether the cause of their existence is merely an artefact of mixing positive and negative item formulations in one scale. From a substantive perspective, we argue that the evidence we have presented in this study points to the possibility that students can hold both fixed and growth mindsets. In other words, the two implicit theories of intelligence are not mutually exclusive but can coexist. We therefore encourage educational researchers to assess both mindsets and represent them as two related but different constructs. Besides, the heterogeneity in the psychometric properties of the ITIS suggests that the sample, study, and measurement contexts matter when interpreting the resultant mindset scores. Our study further demonstrated the utility of item-level meta-analysis for reliability and validity generalization and for testing test theories about educational measurements and their applications.

CRediT authorship contribution statement

Ronny Scherer: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Diego G. Campos:** Conceptualization, Investigation, Data curation, Methodology, Writing – review & editing.

Data availability

Data will be made available on request.

Acknowledgement

Appendix B. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.edurev.2022.100479.

Appendix A

Table A.1
Model Fit Indices and Comparisons Resulting from the Confirmatory Factor Analyses of the Subgroups

Model	χ^2			RMSEA		SRMR	CFI	AIC	BIC	$\Delta\chi^2$		
	Value	df	p	Value	95% CI					Value	Δdf	p
Six-Item ITIS												
Original six-point response scale (N = 12071, m = 24, l = 197)												
Single-factor model	221.9	9	<.001	.044	[.039, .049]	.116	.944	203.9	137.4	–	–	–
Two-factor model	2.7	8	.95	.000	[.000, .000]	.012	1.000	−13.3	−72.5	219.2	1	<.001
Other response scales (N = 15257, m = 15, l = 112)												
Single-factor model	55.9	9	<.001	.019	[.014, .023]	.141	.965	37.9	−30.8	–	–	–
Two-factor model	1.5	8	.99	.000	[.000, .000]	.023	1.000	−14.5	−75.6	54.4	1	<.001
School samples (N = 19363, m = 24, l = 185)												
Single-factor model	300.0	9	<.001	.041	[.037, .045]	.153	.925	282.0	211.2	–	–	–

(*continued on next page*)

Table A.1 (*continued*)

Model	χ²			RMSEA		SRMR	CFI	AIC	BIC	Δχ²		
	Value	df	p	Value	95% CI					Value	Δdf	p
Two-factor model	5.5	8	.71	.000	[.000, .006]	.017	1.000	−10.5	−73.5	294.6	1	<.001
University samples# (N = 7965, m = 15, l = 124)												
Single-factor model	782.5	9	<.001	.104	[.098, .110]	.143	.981	764.5	701.6	–	–	–
Two-factor model	61.1	8	<.001	.029	[.022, .036]	.017	.999	45.1	−10.8	721.4	1	<.001
Online assessment (N = 7408, m = 15, l = 122)												
Single-factor model	130.2	9	<.001	.043	[.036, .049]	.100	.976	112.2	50.0	–	–	–
Two-factor model	8.5	8	.39	.003	[.000, .014]	.018	1.000	−7.5	−62.8	121.7	1	<.001
Paper-and-pencil assessment (N = 19920, m = 24, l = 187)												
Single-factor model	126.0	9	<.001	.026	[.022, .030]	.140	.950	108.0	36.9	–	–	–
Two-factor model	1.5	8	.99	.000	[.000, .000]	.014	1.000	−14.5	−77.7	124.5	1	<.001
Original scales without modifications (N = 15222, m = 26, l = 215)												
Single-factor model	136.3	9	<.001	.031	[.026, .035]	.116	.968	118.3	49.7	–	–	–
Two-factor model	2.4	8	.97	.000	[.000, .000]	.015	1.000	−13.6	−74.7	134.0	1	<.001
Modified scales# (N = 12106, m = 13, l = 94)												
Single-factor model	2497.1	9	<.001	.151	[.146, .156]	.201	.864	2479.1	2412.5	–	–	–
Two-factor model	28.4	8	<.001	.015	[.009, .021]	.017	.999	12.4	−46.8	2468.7	1	<.001
East-Asian samples# (N = 14067, m = 12, l = 58)												
Single-factor model	1739.5	9	<.001	.117	[.112, .122]	.202	.932	1721.5	1653.5	–	–	–
Two-factor model	103.5	8	<.001	.029	[.024, .034]	.076	.996	87.5	27.1	1635.9	1	<.001
Other samples (N = 13261, m = 27, l = 251)												
Single-factor model	170.1	9	<.001	.037	[.032, .042]	.116	.955	152.1	84.7	–	–	–
Two-factor model	0.9	8	.999	.000	[.000, .000]	.008	1.000	−15.1	−75.0	169.2	1	<.001
Eight-Item ITIS												
Original six-point response scale (N = 12071, m = 24, l = 316)												
Single-factor model	409.4	20	<.001	.040	[.037, .044]	.117	.941	369.4	221.5	–	–	–
Two-factor model	11.0	19	.92	.000	[.000, .003]	.019	1.000	−27.0	−167.6	398.4	1	<.001
Other response scales (N = 15257, m = 15, l = 171)												
Single-factor model	72.9	20	<.001	.013	[.010, .017]	.119	.974	32.9	−119.8	–	–	–
Two-factor model	4.6	19	.999	.000	[.000, .000]	.034	1.000	−33.4	−178.5	68.3	1	<.001
School samples (N = 19363, m = 24, l = 263)												
Single-factor model	423.3	20	<.001	.032	[.030, .035]	.158	.930	383.3	225.9	–	–	–
Two-factor model	8.9	19	.98	.000	[.000, .000]	.016	1.000	−29.1	−178.7	414.4	1	<.001
University samples# (N = 7965, m = 15, l = 224)												
Single-factor model	981.7	20	<.001	.078	[.074, .082]	.142	.981	941.7	802.1	–	–	–
Two-factor model	197.8	19	<.001	.034	[.030, .039]	.026	.996	159.8	27.1	783.9	1	<.001
Online assessment (N = 7408, m = 15, l = 212)												
Single-factor model	279.0	20	<.001	.042	[.038, .046]	.101	.964	239.0	100.8	–	–	–
Two-factor model	18.0	19	.52	.000	[.000, .010]	.021	1.000	−20.0	−151.3	261.0	1	<.001
Paper-and-pencil assessment (N = 19920, m = 24, l = 275)												
Single-factor model	175.8	20	<.001	.020	[.017, .023]	.127	.953	135.8	−22.2	–	–	–
Two-factor model	5.5	19	.999	.000	[.000, .000]	.028	1.000	−32.5	−182.6	170.3	1	<.001
Original scales without modifications (N = 15222, m = 26, l = 360)												
Single-factor model	229.7	20	<.001	.026	[.023, .029]	.116	.969	189.7	37.1	–	–	–
Two-factor model	5.4	19	.999	.000	[.000, .000]	.017	1.000	−32.6	−177.6	224.3	1	<.001
Modified scales# (N = 12106, m = 13, l = 127)												
Single-factor model	2567.0	20	<.001	.103	[.099, .106]	.170	.866	2527.0	2379.0	–	–	–
Two-factor model	159.5	19	<.001	.025	[.021, .028]	.069	.993	121.5	−19.2	2407.6	1	<.001
East-Asian samples# (N = 14067, m = 12, l = 80)												
Single-factor model	1830.5	20	<.001	.080	[.077, .083]	.216	.931	1790.5	1639.5	–	–	–
Two-factor model	166.3	19	<.001	.024	[.020, .027]	.068	.994	128.3	−15.2	1664.3	1	<.001
Samples of non-East-Asian origin (N = 13261, m = 27, l = 407)												
Single-factor model	287.2	20	<.001	.032	[.029, .035]	.113	.957	247.2	97.3	–	–	–
Two-factor model	5.3	19	.999	.000	[.000, .000]	.017	1.000	−32.7	−175.1	281.9	1	<.001

Note. RMSEA = Root Mean Square Error of Approximation, *df* = degrees of freedom, 95% CI = 95% confidence interval, SRMR = Standardized Root Mean Squared Residual, CFI = Comparative Fit Index, AIC = Akaike's Information Criterion, BIC = Bayesian Information Criterion, *N* = Overall size of the study samples, *m* = Number of independent samples, *l* = Number of correlation coefficients, ITIS = Implicit Theories of Intelligence Scale. # Given the limited number of samples within this subgroup or unstable model parameter estimates, the stage-1 pooling of the correlation matrices was based on a multivariate fixed-effects model. To ensure the quality of these models, we considered them only if the fit indices suggested at least a reasonable model fit (see Supplementary Material S4 and S5).

Table A.2
Factor Correlations and Scale Reliability Coefficients for the Subgroups

Sample	Six-Item ITIS			Eight-Item ITIS		
Estimates [95% LBCI]	Factor correlation ρ	Reliability ω_F	Reliability ω_G	Factor correlation ρ	Reliability ω_F	Reliability ω_G
Original six-point response scale	.629 [.589, .670]	.839 [.822, .856]	.825 [.802, .847]	.639 [.609, .670]	.872 [.860, .883]	.867 [.855, .878]
Other response scales	.623 [.535, .716]	.826 [.795, .855]	.844 [.812, .874]	.655 [.582, .731]	.862 [.838, .884]	.870 [.849, .888]
School samples	.519 [.472, .568]	.806 [.787, .824]	.785 [.766, .803]	.512 [.471, .554]	.842 [.829, .856]	.834 [.821, .847]
University samples	.770 [.753, .787]	.895 [.891, .899]	.901 [.895, .906]	.802 [.788, .816]	.927 [.924, .930]	.915 [.911, .919]
Online assessment	.730 [.690, .771]	.837 [.812, .861]	.853 [.840, .865]	.720 [.691, .749]	.874 [.860, .888]	.884 [.875, .893]
Paper-and-pencil assessment	.576 [.514, .641]	.835 [.816, .854]	.819 [.786, .850]	.602 [.550, .655]	.866 [.849, .882]	.861 [.841, .879]
Original scales without modifications	.681 [.633, .730]	.845 [.828, .861]	.858 [.838, .877]	.685 [.647, .723]	.880 [.869, .890]	.885 [.874, .896]
Modified scales	.418 [.396, .440]	.777 [.769, .784]	.801 [.791, .810]	.440 [.419, .462]	.811 [.799, .822]	.825 [.815, .835]
East-Asian samples	.476 [.451, .501]	.821 [.816, .826]	.792 [.780, .804]	.479 [.455, .504]	.855 [.849, .862]	.849 [.837, .859]
Samples of non-East-Asian origin	.649 [.603, .696]	.832 [.815, .849]	.835 [.814, .856]	.660 [.624, .696]	.871 [.858, .882]	.868 [.856, .879]

Note. ITIS = Implicit Theories of Intelligence Scale, 95% LBCI = 95% Likelihood-based confidence interval, ω_F = Reliability coefficient of the fixed-mindset subscale, ω_G = Reliability coefficient of the growth-mindset subscale. The reliability coefficient ω refers to the scale reliability $\omega = (\sum\limits_{i=1}^{I} \lambda_i)^2 / [(\sum\limits_{i=1}^{I} \lambda_i)^2 + \sum\limits_{i=1}^{I} (1 - \lambda_i^2)]$ based on items $i = 1, \ldots, I$ with factor loadings λ_i and residual variances $1 - \lambda_i^2$ (Scherer & Teo, 2020).

References

Arens, A. K., Jansen, M., Preckel, F., Schmidt, I., & Brunner, M. (2020). The structure of academic self-concept: A methodological review and empirical illustration of central models. *Review of Educational Research, 91*(1), 34–72. https://doi.org/10.3102/0034654320972186

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110*(2), 203–219. https://doi.org/10.1037/0033-295X.110.2.203

Bostwick, K. C. P., Collie, R. J., Martin, A. J., & Durksen, T. L. (2017). Students' growth mindsets, goals, and academic outcomes in mathematics. *Zeitschrift für Psychologie, 225*(2), 107–116. https://doi.org/10.1027/2151-2604/a000287

Braeken, J., & van Assen, M. A. L. M. (2017). An empirical Kaiser criterion. *Psychological Methods, 22*(3), 450–466. https://doi.org/10.1037/met0000074

Burgoyne, A. P., & Macnamara, B. N. (2021). Reconsidering the use of the mindset assessment profile in educational contexts. *Journal of Intelligence, 9*(3), 39. https://doi.org/10.3390/jintelligence9030039

Burnette, J. L., O'Boyle, E. H., VanEpps, E. M., Pollack, J. M., & Finkel, E. J. (2013). Mind-sets matter: A meta-analytic review of implicit theories and self-regulation. *Psychological Bulletin, 139*(3), 655–701. https://doi.org/10.1037/a0029531

Carpenter, N. C., Son, J., Harris, T. B., Alexander, A. L., & Horner, M. T. (2016). Don't forget the items: Item-level meta-analytic and substantive validity techniques for reexamining scale validation. *Organizational Research Methods, 19*(4), 616–650. https://doi.org/10.1177/1094428116639132

Cheung, M. W.-L. (2013). Multivariate meta-analysis as structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 20*(3), 429–454. https://doi.org/10.1080/10705511.2013.797827

Cheung, M. W.-L. (2015a). *Meta-analysis: A structural equation modeling approach.* John Wiley & Sons Ltd.

Cheung, M. W.-L. (2015b). metaSEM: An R package for meta-analysis using structural equation modeling. *Frontiers in Psychology, 5*(1521). https://doi.org/10.3389/fpsyg.2014.01521

Claro, S., Paunesku, D., & Dweck, C. S. (2016). Growth mindset tempers the effects of poverty on academic achievement. *Proceedings of the National Academy of Sciences, 113*(31), 8664–8668. https://doi.org/10.1073/pnas.1608207113

Cook, D. A., Castillo, R. M., Gas, B., & Artino, A. R., Jr. (2017). Measuring achievement goal motivation, mindsets and cognitive load: Validation of three instruments' scores. *Medical Education, 51*(10), 1061–1074. https://doi.org/10.1111/medu.13405

Costa, A., & Faria, L. (2018). Implicit theories of intelligence and academic achievement: A meta-analytic review. *Frontiers in Psychology, 9*(829). https://doi.org/10.3389/fpsyg.2018.00829

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods, 14*(2), 81–100. https://doi.org/10.1037/a0015914

Diseth, Å., Meland, E., & Breidablik, H. J. (2014). Self-beliefs among students: Grade level and gender differences in self-esteem, self-efficacy and implicit theories of intelligence. *Learning and Individual Differences, 35*, 1–8. https://doi.org/10.1016/j.lindif.2014.06.003

Dweck, C. S. (2000). *Self-Theories: Their role in motivation, personality, and development.* Psychology Press.

Dweck, C. S., Chiu, C.-y., & Hong, Y.-y. (1995). Implicit theories and their role in judgments and reactions: A world from two perspectives. *Psychological Inquiry, 6*(4), 267–285. https://doi.org/10.1207/s15327965pli0604_1

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*(2), 256–273. https://doi.org/10.1037/0033-295X.95.2.256

Dweck, C. S., & Yeager, D. S. (2019). Mindsets: A view from two eras. *Perspectives on Psychological Science, 14*(3), 481–496. https://doi.org/10.1177/1745691618804166

Epskamp, S. (2021). psychonetrics: Structural equation modeling and confirmatory network analysis. *R package version 0.9* https://CRAN.R-project.org/package=psychonetrics.

Epskamp, S., Isvoranu, A.-M., & Cheung, M. W. L. (2022). Meta-analytic Gaussian network aggregation. *Psychometrika, 87*, 12–46. https://doi.org/10.1007/s11336-021-09764-3

Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika, 82*(4), 904–927. https://doi.org/10.1007/s11336-017-9557-x

Fisher, Z., Tipton, E., & Zhipeng, H. (2017). *Robumeta: Robust variance meta-regression. R package version 2.0* https://CRAN.R-project.org/package=robumeta.

Freund, P. A., & Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychological Bulletin, 138*(2), 296–321. https://doi.org/10.1037/a0026556

Glerum, J., Loyens, S. M. M., & Rikers, R. M. J. P. (2020). Mind your mindset. An empirical study of mindset in secondary vocational education and training. *Educational Studies, 46*(3), 273–281. https://doi.org/10.1080/03055698.2019.1573658

Gnambs, T., & Schroeders, U. (2017). Cognitive abilities explain wording effects in the Rosenberg self-esteem scale. *Assessment, 27*(2), 404–418. https://doi.org/10.1177/1073191117746503

Gonzalez, O., MacKinnon, D. P., & Muniz, F. B. (2021). Extrinsic convergent validity evidence to prevent Jingle and Jangle fallacies. *Multivariate Behavioral Research, 56*(1), 3–19. https://doi.org/10.1080/00273171.2019.1707061

Hass, R. W., Reiter-Palmon, R., & Katz-Buonincontro, J. (2017). Chapter 12 - are implicit theories of creativity domain specific? Evidence and implications. In M. Karwowski, & J. C. Kaufman (Eds.), *The creative self* (pp. 219–234). Academic Press. https://doi.org/10.1016/B978-0-12-809790-8.00012-1.

Hedges, L. V., & Schauer, J. M. (2019). Statistical analyses for studying replication: Meta-analytic perspectives. *Psychological Methods, 24*(5), 557–570. https://doi.org/10.1037/met0000189

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions*. Wiley. https://doi.org/10.1002/9781119536604

Hu, L.t., & 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

Jak, S., & Cheung, M. W.-L. (2020). Meta-analytic structural equation modeling with moderating effects on SEM parameters. *Psychological Methods, 25*(4), 430–455. https://doi.org/10.1037/met0000245

Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark Iv. *Educational and Psychological Measurement, 34*(1), 111–117. https://doi.org/10.1177/001316447403400115

Kam, C. C. S. (2016). Why do we still have an impoverished understanding of the item wording effect? An empirical examination. *Sociological Methods & Research, 47*(3), 574–597. https://doi.org/10.1177/0049124115626177

Kan, K.-J., van der Maas, H. L. J., & Levine, S. Z. (2019). Extending psychometric network analysis: Empirical evidence against g in favor of mutualism? *Intelligence, 73*, 52–62. https://doi.org/10.1016/j.intell.2018.12.004

Lewis, K. M., Donnellan, M. B., Ribeiro, J. S., & Trzesniewski, K. (2021). Evaluating evidence for a global mindset factor across multiple ability domains. *Journal of Research in Personality, 95*, Article 104165. https://doi.org/10.1016/j.jrp.2021.104165

Li, Y., & Bates, T. C. (2020). Testing the association of growth mindset and grades across a challenging transition: Is growth mindset associated with grades? *Intelligence, 81*, Article 101471. https://doi.org/10.1016/j.intell.2020.101471

Li, Y., Wen, Z., Hau, K.-T., Yuan, K.-H., & Peng, Y. (2020). Effects of cross-loadings on determining the number of factors to retain. *Structural Equation Modeling: A Multidisciplinary Journal, 27*(6), 841–863. https://doi.org/10.1080/10705511.2020.1745075

Logan, J. A. R., Hart, S. A., & Schatschneider, C. (2021). Data sharing in education science. *AERA Open, 7*. https://doi.org/10.1177/23328584211006475

López-López, J. A., Page, M. J., Lipsey, M. W., & Higgins, J. P. T. (2018). Dealing with effect size multiplicity in systematic reviews and meta-analyses. *Research Synthesis Methods, 9*(3), 336–351. https://doi.org/10.1002/jrsm.1310

Lou, N. M., Chaffee, K. E., & Noels, K. A. (2021). *Growth, fixed, and mixed mindsets: Mindset system profiles in foreign language learners and their role in engagement and achievement. Studies in Second Language Acquisition* (pp. 1–26). https://doi.org/10.1017/S0272263121000401

Lüftenegger, M., & Chen, J. A. (2017). Conceptual issues and assessment of implicit theories. *Zeitschrift für Psychologie, 225*(2), 99–106. https://doi.org/10.1027/2151-2604/a000286

Macnamara, B. N., & Rupani, N. S. (2017). The relationship between intelligence and mindset. *Intelligence, 64*, 52–59. https://doi.org/10.1016/j.intell.2017.07.003

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2

McNeish, D., & Wolf, M. G. (2021). Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychological Methods*. https://doi.org/10.1037/met0000425

Muradoglu, M., Porter, T., Trzesniewski, K., & Cimpian, A. (2022). GM-C: A growth mindset scale for young children. *PsyArXiv Preprints* https://doi.org/10.31234/osf.io/fgw8t.

OECD. (2021). *PISA 2018: Sky's the limit—growth mindset, students, and schools in PISA*. OECD Publishing. https://www.oecd.org/pisa/growth-mindset.pdf.

Paunesku, D., Walton, G. M., Romero, C., Smith, E. N., Yeager, D. S., & Dweck, C. S. (2015). Mind-set interventions are a scalable treatment for academic underachievement. *Psychological Science, 26*(6), 784–793. https://doi.org/10.1177/0956797615571017

Pellegrino, J. W., DiBello, L. V., & Goldman, S. R. (2016). A framework for conceptualizing and evaluating the validity of instructionally relevant assessments. *Educational Psychologist, 51*(1), 59–81. https://doi.org/10.1080/00461520.2016.1145550

Petscher, Y., Al Otaiba, S., Wanzek, J., Rivas, B., & Jones, F. (2017). The relation between global and specific mindset with reading outcomes for elementary school students. *Scientific Studies of Reading, 21*(5), 376–391. https://doi.org/10.1080/10888438.2017.1313846

Pigott, T. D., & Polanin, J. R. (2020). Methodological guidance paper: High-quality meta-analysis in a systematic review. *Review of Educational Research, 90*(1), 24–46. https://doi.org/10.3102/0034654319877153

Plucker, J. A., & Makel, M. C. (2021). Replication is important for educational psychology: Recent developments and key issues. *Educational Psychologist, 56*(2), 90–100. https://doi.org/10.1080/00461520.2021.1895796

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Rammstedt, B., Grüning, D. J., & Lechner, C. M. (2021). Measuring growth mindset: A validation of a three-item scale and a single-item scale in youth and adults. *PsyArXiv Preprints*. https://doi.org/10.31234/osf.io/rs43g

Sarrasin, J. B., Nenciovici, L., Foisy, L.-M. B., Allaire-Duquette, G., Riopel, M., & Masson, S. (2018). Effects of teaching the concept of neuroplasticity to induce a growth mindset on motivation, achievement, and brain activity: A meta-analysis. *Trends in Neuroscience and Education, 12*, 22–31. https://doi.org/10.1016/j.tine.2018.07.003

Scherer, R., & Teo, T. (2020). A tutorial on the meta-analytic structural equation modeling of reliability coefficients. *Psychological Methods, 25*(6), 747–775. https://doi.org/10.1037/met0000261

van de Schoot, R., de Bruin, J., Schram, R., Zahedi, P., de Boer, J., Weijdema, F., Kramer, B., Huijts, M., Hoogerwerf, M., Ferdinands, G., Harkema, A., Willemsen, J., Ma, Y., Fang, Q., Hindriks, S., Tummers, L., & Oberski, D. L. (2021). An open source machine learning framework for efficient and transparent systematic reviews. *Nature Machine Intelligence, 3*(2), 125–133. https://doi.org/10.1038/s42256-020-00287-7

Schroeders, U., Schipolowski, S., & Wilhelm, O. (2015). Age-related changes in the mean and covariance structure of fluid and crystallized intelligence in childhood and adolescence. *Intelligence, 48*, 15–29. https://doi.org/10.1016/j.intell.2014.10.006

Schunk, D. H. (1995). Implicit theories and achievement behavior. *Psychological Inquiry, 6*(4), 311–314. https://doi.org/10.1207/s15327965pli0604_9

Sigmundsson, H., Guðnason, S., & Jóhannsdóttir, S. (2021). Passion, grit and mindset: Exploring gender differences. *New Ideas in Psychology, 63*, 100878. https://doi.org/10.1016/j.newideapsych.2021.100878

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). p-Curve and effect size:correcting for publication bias using only significant results. *Perspectives on Psychological Science, 9*(6), 666–681. https://doi.org/10.1177/1745691614553988

Sisk, V. F., Burgoyne, A. P., Sun, J., Butler, J. L., & Macnamara, B. N. (2018). To what extent and under which circumstances are growth mind-sets important to academic achievement? Two meta-analyses. *Psychological Science, 29*(4), 549–571. https://doi.org/10.1177/0956797617739704

Steinmann, I., Strietholt, R., & Braeken, J. (2021). A constrained factor mixture analysis model for consistent and inconsistent respondents to mixed-worded scales. *Psychological Methods*. https://doi.org/10.1037/met0000392

Sternberg, R. J. (1985). Implicit theories of intelligence, creativity, and wisdom. *Journal of Personality and Social Psychology, 49*(3), 607–627. https://doi.org/10.1037/0022-3514.49.3.607

Tempelaar, D. T., Rienties, B., Giesbers, B., & Gijselaers, W. H. (2015). The pivotal role of effort beliefs in mediating implicit theories of intelligence and achievement goals and academic motivations. *Social Psychology of Education, 18*(1), 101–120. https://doi.org/10.1007/s11218-014-9281-7

UNESCO Institute for Statistics. (2012). *International standard classification of education ISCED 2011*. UNESCO Institute for Statistics. http://uis.unesco.org/en/topic/international-standard-classification-education-isced.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3). https://www.jstatsoft.org/v36/i03/.

Viechtbauer, W., & Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods, 1*(2), 112–125. https://doi.org/10.1002/jrsm.11

Wang, Q., & Ng, F. F.-Y. (2012). Chinese students' implicit theories of intelligence and school performance: Implications for their approach to schoolwork. *Personality and Individual Differences, 52*(8), 930–935. https://doi.org/10.1016/j.paid.2012.01.024

Yeager, D. S., & Dweck, C. S. (2020). What can be learned from growth mindset controversies? *American Psychologist, 75*(9), 1269–1284. https://doi.org/10.1037/amp0000794

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., Tipton, E., Schneider, B., Hulleman, C. S., Hinojosa, C. P., Paunesku, D., Romero, C., Flint, K., Roberts, A., Trott, J., Iachan, R., Buontempo, J., Yang, S. M., Carvalho, C. M., … Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature, 573*(7774), 364–369. https://doi.org/10.1038/s41586-019-1466-y