

Development and Validation of the “Tübingen Inventory to Measure Teachers’ Profession-Specific Value Orientations” (TIVO)

Samuel Merk 

University of Education Karlsruhe

Martin Drahmann

Colin Cramer 

University of Tübingen

Broadly understood, values represent orientation guidelines for daily action, thinking, and feeling. Thus, they affect teachers’ everyday work and are among their professional competencies. While general values and some single profession-specific values of teachers (e.g., responsibility) have already been investigated empirically, the current study aims to cover a broader range of profession-specific values by developing and validating the Tübingen Inventory to Measure Teachers’ Profession-Specific Value Orientations (TIVO), based on three independent studies with pre-service (Studies 1 and 2; $N_1 = 334$, $N_2 = 239$) and in-service teachers ($N_3 = 308$). The results demonstrate that the TIVO is appropriate to assess four profession-specific values in a second-order model: caring, justice, responsibility, and truthfulness as first-order factors, with fairness as a second-order factor loading on the latter three first-order factors. The results from preregistered experiments and confirmatory factor analysis provide consistent evidence for the construct validity of the TIVO.

Keywords: *teachers’ values, assessing values, teacher education, moral dimension of teaching*

SEVERAL disciplines, including philosophy, psychology, and sociology, place great importance on values (Krobath, 2009). For instance, values have been cited as the basis for explaining behavior patterns, attitudes, or motives for and goals of action (Kluckhohn, 1951; Schwartz, 1992); even social subsystems can be distinguished in terms of their value preferences (Bond, 1988; Höffe, 2008; Schwartz, 2011). The importance of values in teachers’ professional action has been particularly emphasized in the literature addressing the moral dimension of teaching (Carr, 2010; Klaassen et al., 2016; Oser, 1994).

Although the empirical investigation of general values in different societies and for different social subsystems (Inglehart et al., 2000; Lindeman & Verkasalo, 2005; Rokeach, 1973; Vernon & Allport, 1931) has spawned a broad literature, studies addressing the profession-specific values of teachers are sparse (Lauermaun & Karabenick, 2013).

The present article aims to validate the Tübingen Inventory to Measure Teachers’ Profession-Specific Value Orientations (TIVO). The first part of the article defines values and highlights current approaches to their assessment, and the second part provides an overview of literature

addressing the profession-specific values of teachers. Finally, we describe how we distilled five profession-specific values and developed and validated the TIVO by leveraging three empirical studies.

General Values

Values are considered comparatively rarely in the literature on education and educational psychology (Fries et al., 2007). Hence, we offer a brief definition and a list of criteria to distinguish them from related constructs. According to Wray-Lake et al. (2014), “values refer to abstract, emotionally valenced, higher-order beliefs that exist along a continuum of importance and guide more specific attitudes and behaviors” (p. 1102). Schwartz (2007) notes that six features emerge as the conceptual basis for a theory of values: (1) they are beliefs and hence personal truth propositions that are connected to affect; (2) they are connected to desirable goals that motivate behavior; (3) they are not specific to contexts, objects, or events, e.g., sports or family; (4) they guide actions by serving as individual standards; (5) individuals exhibit relative importance of different values; and (6) as



most behavior has implications for multiple values, this *relative* importance guides individuals' actions through their consideration of trade-offs.

The first defining criterion can be used to distinguish between goals and values. Whereas goals are related to desirable end states, values are state-independent propositions, although both have motivational power. Attitudes share with (human) values that they assess entities, but attitudes do so with respect to very concrete events and objects and along a continuum of approval and disapproval. The four subcategories of task values from expectancy value theory (Eccles et al., 1983; Eccles & Wigfield, 2002) exist along a continuum of importance but refer to concrete and context-specific entities—the tasks—whereas (human) values transcend specific actions and situations. Finally, one can distinguish norms from values using the fourth defining characteristic: Norms are about oughtness (Marini, 2000), which is defined beyond the individual whose own values serve as standards for actions.

Another conceptual difficulty is the subtle differentiation between values and value orientations. Some authors use the term *value* to address abstract entities (like “authority”) to which the beliefs of individuals refer and define these cognitively represented beliefs as “value orientations” (e.g., Zhu & Chen, 2018). Others define “values” as the beliefs (see above) and “value orientations” as the central values emphasized in a society or social subgroup (e.g., Schwartz, 2007), while still others use “values” and “value orientations” more or less as synonyms for individual beliefs (e.g., Heim et al., 2017). A confusion of the level of values and value orientations may also be promoted by the fact that classic instruments like the Portrait Values Questionnaire (Schwartz & Cieciuch, 2021) or the Rokeach Value Survey (Rokeach, 1973) are used to investigate both individuals' values and cultural value orientations.

Teachers' (Profession-Specific) Values

In recent decades, several researchers from different disciplines have emphasized the general importance of values in the teaching profession (Carr, 2010; Clayman, 1961; Gudmundsdottir, 1990; James & McCormick, 2009; Wannamaker & Tennyson, 1970). This attribution of importance is visible in often-cited topological models of teacher knowledge or teachers' professional competencies (Baumert & Kunter 2006; Shulman, 1987). Referring to Shulman's model, Gudmundsdottir (1990) points out that excellent teachers' “values are an integral part of their excellence in teaching” (p. 50), while Baumert and Kunter (2006) designate an extra aspect of professional competence to the beliefs, values, and goals of teachers, alongside more prominent aspects like subject knowledge and general pedagogical knowledge.

There are three strands in the literature about teachers' profession-specific values. One is research on teacher ethos,

which largely nonempirically addresses the moral dimension of teaching. This strand focuses theoretically on the ethical dilemmas of the teaching profession (Nash, 1991) and the role of teachers' values in their competency (Oser, 1994). A second strand comprises empirical studies that focus on selected profession-specific values. Lauermaann and Karabenick (2013), for example, focus on teachers' responsibility, which they define “as a sense of internal obligation and commitment to produce or prevent designated outcomes, or that these outcomes should have been produced or prevented” (p. 13). Their empirical investigation provided evidence for the assumption that responsibility can be seen as a multidimensional construct that encompasses responsibility for student motivation, student achievement, relationships with students, and teaching, all of which have divergent validities as to self-efficacy. As a third strand, we identified studies that target the profession specificity of teachers' values by between-person studies using instruments assessing general values. For example, Mägdefrau (2008) compared student teachers with business and engineering students using a general value survey. She found teacher students to be more socially oriented and more conservative in their general values.

It should be noted that several studies address teachers' “valuing” of specific pedagogical tasks. For example, James and McCormick (2009) had teachers rate how important they thought it was to make learning explicit, promote learning autonomy, and adopt a performance orientation. However, as task-specific values are conceptually very different from teachers' profession-specific values (see the definition above), these studies are not within the scope of this article.

Current Studies

As illustrated above, the current literature about teachers' profession-specific values highlights their importance, but empirical work mostly focuses on a single value (e.g., responsibility) or on differences in general values between teachers and nonteachers. The aim of the present work is to cover a broader range of profession-specific values by developing and validating the TIVO. Therefore, we first identified five profession-specific values from the literature (see the following section) and then focused on factorial validity (Study 1: exploratory factor analyses [EFAs]; Studies 2 and 3: confirmatory factor analyses [CFAs]) and experimental construct validation (Studies 1 and 2).

Identification of Profession-Specific Values

To broaden the empirical assessment of teachers' profession-specific values, we skimmed the extensive literature of teacher ethos and the moral dimension of teaching (Oser, 1994; Oser & Biedermann, 2018). This literature has

proposed a number of profession-specific values for the teaching profession as focal (e.g., Carr, 2006, 2010; Oser, 1998; Zia, 2007), relying mostly on nonempirical methods. Therefore, the authors created a list of 25 articles that they subjectively determined were the most promising for this endeavor (for a complete list, see the reproducible documentation of analysis [RDA] at the Open Science Framework; <https://osf.io/bqmw9>). A content analysis (Krippendorff, 2019) of this literature revealed that the five most frequently appearing values were caring (mentioned in 20 articles), justice (14), truthfulness (11), tolerance (10), and responsibility (9).

Caring (e.g., Noddings, 1984) focuses on the relationship between teachers and students. In this relationship (which needs to be constantly established and cultivated), the mutual appreciation of and the respect for the other are crucial elements of a formative process (Thayer-Bacon, 2008). Justice (e.g., Kohlberg, 1981) indicates that all students are treated equally by their teachers according to their individual requirements or achievements. Equality or reciprocity are important, because it is not the needs of the individual that are addressed but the regulation of the claims to validity of different positions (Oser, 1998). Truthfulness (e.g., Veugelers, 2010) is expressed when teachers' opinions are determined neither by an overemphasis on caring or justice nor any other external expectation. Decisions have to be justified and must be made in accordance with one's own values—disagreement must be handled faithfully and in a cooperative manner to achieve consensus in the classroom (Oser, 1998). Tolerance (e.g., Horton, 1998, p. 429) manifests itself by the conscious decision not to prohibit, impede, or take action against disapproving behaviors, even though one would have the position, right, or opportunity to do so. It expresses itself in teachers' acceptance and understanding of other people's opinions and characteristics or in a universalistic attitude toward all people and belongings (e.g., Harder, 2014). Responsibility (e.g., Weinberger et al., 2018) refers to a teacher's obligation to ensure that the fulfillment of a task takes the best possible course and that no damage occurs. Acting responsibly as a teacher means taking responsibility for someone (e.g., students in class) and taking responsibility for something (e.g., given legal bases). The sixth most frequently mentioned value was fairness. As it occurred only five times and was much more vaguely defined, we decided not to consider this or any other values, as the aim was not to exhaustively assess profession-specific values.

We then skimmed these 25 articles again to identify adjectives used to describe these five dimensions, because we planned to construct the TIVO as a semantic differential. A content validation of these adjectives then took place. $N = 14$ experts rated the content validity of each adjective and were asked to suggest appropriate antonyms for each adjective. Finally, two additional experts were

cognitively interviewed (Willis, 2005) regarding their deliberations on the relation between each adjective and its respective dimension. As a result of this process, we identified 40 adjective pairs (eight per dimension; see RDA).

Developing Experimental Materials for Construct Validation (Vignettes)

To experimentally investigate the construct validity (Cronbach & Meehl, 1955; Messick, 1995), the authors created 10 text vignettes (two per value); each one began with an everyday situation in which one of the five values became prominent (see Table 1 for an example). The text then proceeds with a description of a teacher with either a very high or very low manifestation of this value. In Studies 1 and 2, participants had to apply the TIVO to the teachers described in the vignettes. Leveraging the concept of construct validity, one would expect stronger differences in the ratings regarding the manipulated dimension.

Study 1 (Exploratory Study)

Design

Study 1 consists of two main parts. First, respondents were encouraged to provide a self-description using the semantic differential, while the second part was experimental in nature. To investigate construct validity (Cronbach & Meehl, 1955; Messick, 1995), we employed the previously developed text vignettes, each of which describes one of the five proposed dimensions of values of teachers in either a high or low specification. Participants were then confronted with the vignettes and prompted to use the semantic differential to describe the teacher in the text vignette. To avoid a contrast effect (Schwarz, 1999), every participant was presented only one version of the two vignettes. Furthermore, each participant was presented with only three vignettes to keep the complete survey economic. The sequence of the vignettes was block-randomized using incomplete Latin Squares.

Sample

The sample for Study 1 was recruited in lectures and courses in teacher education at a university in Germany. It consists of $N = 334$ student teachers (216 female, $M_{\text{semester}} = 4.15$, $SD_{\text{semester}} = 1.03$), whose participation was voluntary and unrewarded. The survey was carried out using paper-and-pencil procedures administered by trained test conductors as a groupwise assessment during academic coursework.

Procedure and Materials

In the first part of the questionnaire, the participating teacher students were prompted to describe themselves

TABLE 1

Example of a Vignette Describing a High or Low Expression of Caring

Joint introduction of both text vignettes

At a staff meeting, teachers are informed about a female student whose performance has been declining drastically over the past few months. Most likely, she will not meet the required standards and will not move on to the next grade. In this context, the principal points to the student's severely difficult situation at home and in her class at the present time. This is followed by a discussion of two teachers in which one of them emphasizes that . . .

High expression of caring

. . . she/he feels compelled to check on the student more often and to seek a dialogue. To increase her performance, the student should be supported beyond the means of subject teaching. Relationship building and the well-being of the student— not just achievement—are of importance.

Low expression of caring

. . . she/he does not feel compelled to check on the student and she/he will bring home the message. Regardless of personal or social problems, subject-specific achievement is paramount. That is why she/he will not offer any ongoing support. Only subject-specific achievement is of importance, not relationship building or the well-being of the student.

using the 40 adjective pairs (“You will now see several pairs of characteristics. Please try to assess your professional behavior as a (future) teacher based on the following pairs of characteristics. Some adjective pairs may not always seem appropriate to you. Nevertheless, please try to make a personal assessment for each pair”). In the second part, participants judged the values of three teachers described in the text vignettes by rating the TIVO for each vignette. In each case, participants were asked first to carefully read the vignette describing a fictitious teacher (see Table 1) and then to rate the teacher's behavior and/or statements using the adjective pairs (see online Supplemental Material).

Results

We first checked the appropriateness of the data using Kaiser–Meyer–Olkin statistics. As the value for the whole sample was .92 and the minimum for the item-specific values was .81, we judged the data to be factorable. Additionally, we computed Bartlett's test, which was highly significant, $\chi^2(780) = 7224.0, p < .001$, and checked item intercorrelations, which were all less than .79.

To determine the number of factors to be extracted, we used scree plots based on maximum-likelihood exploratory factor analysis (ML-EFA), the very simple structure (VSS) criterion (Revelle & Rocklin, 1979), the empirical Bayesian information criterion (BIC), and parallel analysis (Horn, 1965). The visual inspection of the scree plots favored a two-factor solution for the self-describing ratings of the TIVO and three of the five text vignette ratings (see RDA). The VSS (with complexity 2) favored two factors for all ratings whereby the BIC achieved a minimum assuming five factors for the self-description and two factors for three vignette ratings, and three factors for the vignette responsibility. Parallel analysis finally suggested six factors for the self-describing ratings, two factors for three vignettes, and

three factors for the vignette responsibility. Focusing on interpretability, we inspected the loading patterns for all suggested solutions carefully. Despite the heterogeneous results regarding the number of factors, the loading patterns consistently revealed a distinction between the adjectives a priori mapped to the dimension caring and the other adjectives. We thus applied ML-EFA for two factors with oblimin rotation to the self-description answers and to the answers to every vignette. The results are presented in Figure 1 (for detailed tables, see RDA).

There, adjectives from the proposed caring dimension are strongly associated with Factor 1, and several items from the proposed truthfulness and responsibility dimensions are associated with Factor 2. However, several items alternate loading on both factors or load on neither. Given the challenge of choosing a final item set in light of these results, we decided to weigh the results from the self-description more heavily and to incorporate thoughts about the content validity of the items. This resulted in a set of 18 items displayed on the right side of Figure 1. It turns out that Factor 1 loads on five of eight adjective pairs concerning caring, so this Factor 1 is labelled “caring.” In addition, this factor loads on two adjective pairs with reference to justice and tolerance, which both semantically show a high similarity to caring and are therefore retained in Factor 1. Factor 2 loads on five pairs of adjectives related to justice and three pairs of adjectives related to both, truthfulness and responsibility. “Fairness” is chosen as the label of this second factor because the three dimensions of justice, truthfulness, and responsibility are theoretically reflected in the construct of fairness (Höffe, 2008).

To explore the degree to which the initially proposed dimensions of truthfulness, responsibility, and justice are separable, we fitted three CFA models on the selected items using the full information maximum-likelihood estimator available within the R package lavaan (Rosseel, 2012). The

Loadings of the initial and selected items

Cut off for loadings = .4

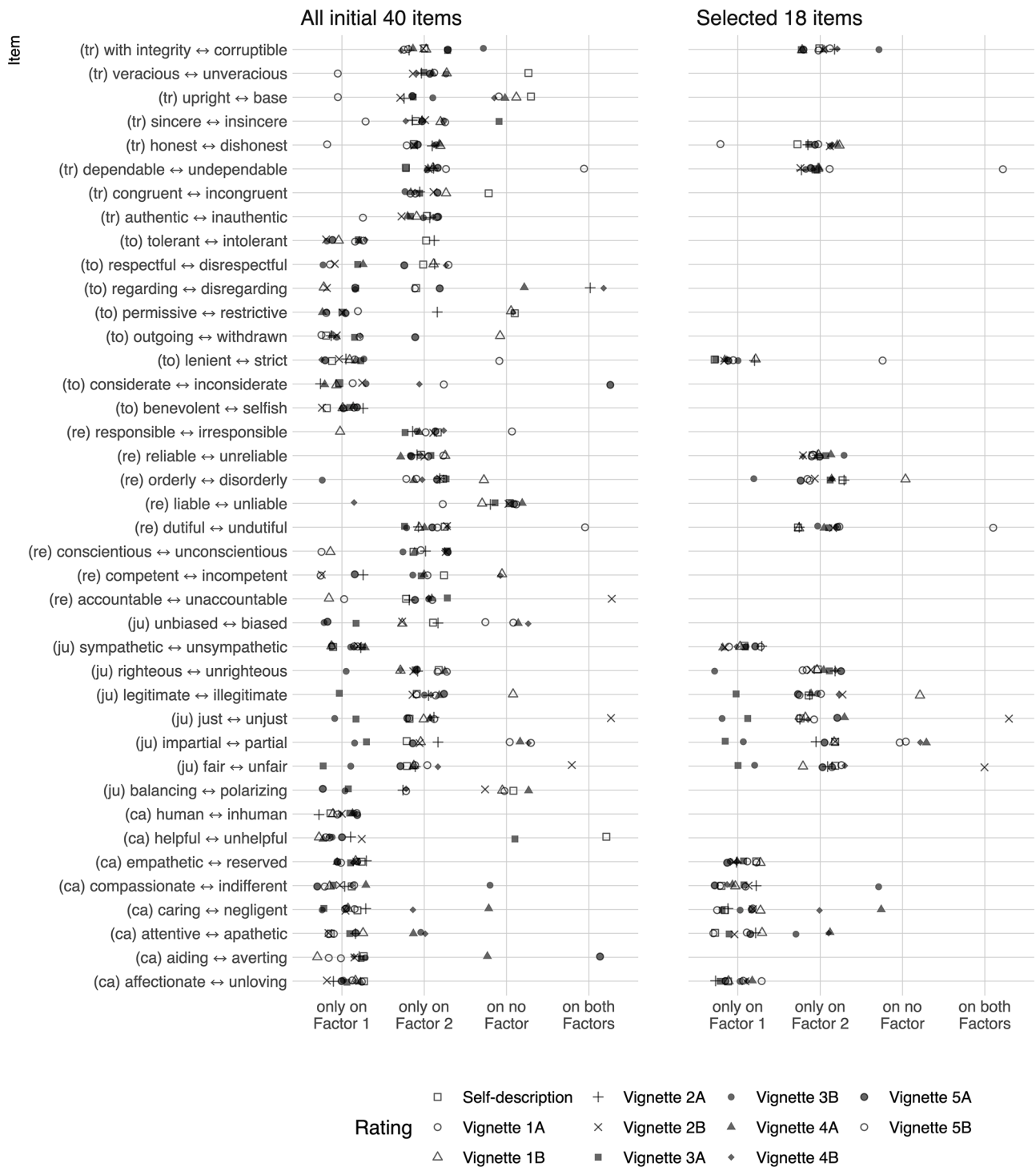


FIGURE 1. Results from maximum-likelihood exploratory factor analysis (ML-EFA).
 Note. ca = caring; ju = justice; to = tolerance; re = responsibility; tr = truthfulness.

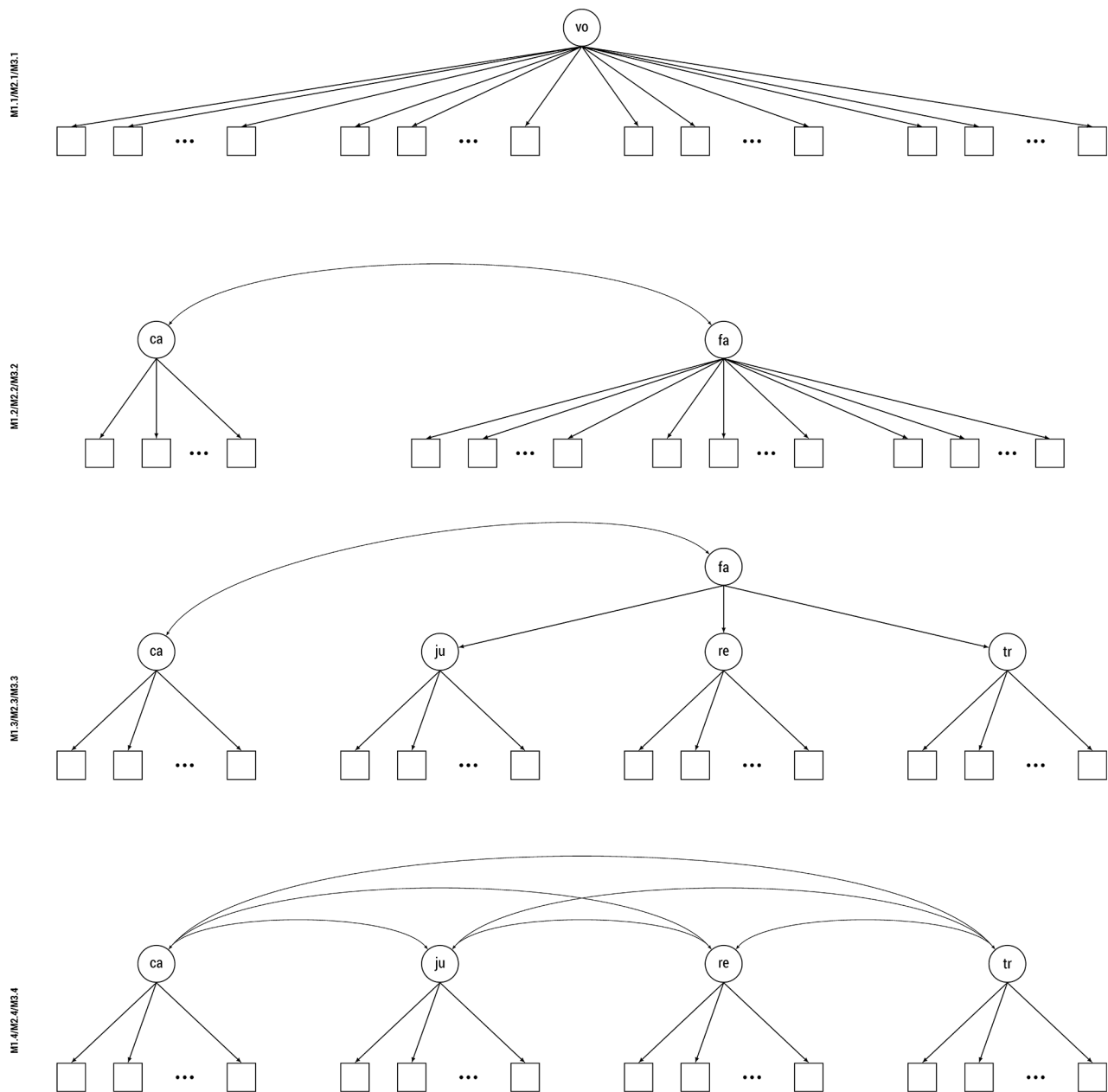


FIGURE 2. Tested confirmatory factor analysis (CFA) models in Studies 1 and 2.

Note. ca = caring; ju = justice; re = responsibility; tr = truthfulness; vo = value orientation; MX.Y = CFA-Model Y in Study X. Residual covariances are not shown.

first model we fitted (M1.2; see Figure 2) as a reference model had two factors and congeneric measurement models analogous to the ML-EFA results. After freeing four residual covariance parameters chosen based on modification indices, this model showed the following model fit: comparative fit index (CFI) = .921, Tucker–Lewis index (TLI) = .907; root mean square error of approximation (RMSEA) = .065; and standardized root mean square residual (SRMR) = .055, which is usually judged as acceptable (Marsh et al., 2004;

Nagengast et al., 2013). The next model (M1.3; see Figure 2) reflected the initial mapping of the remaining items and the ML-EFA results using a second-order structure (Brown, 2015). M1.3 also showed good model fit (CFI = .927, TLI = .912, RMSEA = .063, SRMR = .053), and a chi-square difference test became significant, indicating significantly better model fit for M1.3 than M1.2 when considering the additional degrees of freedom in M1.3. In a final model (M1.4; see Figure 2), we specified four factors along the

Means \pm 1SD of Caring and Fairness

Subplots indicate manipulated dimensions

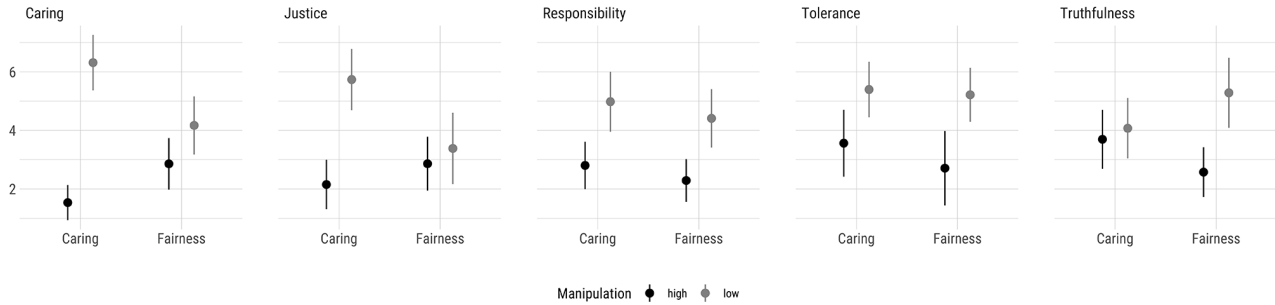


FIGURE 3. *Effects of the manipulation (Study 1).*

Note. Means and standard deviations were computed from mean scores of the fairness and caring scales.

initial mapping of the remaining items. This model also showed good model fit (CFI = .926, TLI = .909, RMSEA = .064, SRMR = .052), but the chi-square difference test (M1.4 vs. M1.3) was not significant (.725). As we specified congeneric measurement models, we used McDonald's ω as a measure of internal consistency (Dunn et al., 2014), with the results indicating strong internal consistency for two first-order factors and the second-order factor (caring ω = .829; justice ω = .728; fairness ω = .956) and weak internal consistency for two first-order factors that are both part of the second order factor (responsibility ω = .614, truthfulness ω = .584).

Part two of Study 1 aimed to find evidence for the construct validity by asking respondents to judge experimentally manipulated descriptions of fictitious teachers using the semantic differential. Figure 2 depicts the scores of the caring and fairness dimensions grouped into subplots by the four initial dimensions, which were manipulated; colors encode the direction of manipulation.

As we did not expect that manipulating one dimension would have no effect on the other dimensions, *larger* effect sizes for the manipulated dimension have been hypothesized than for those not manipulated. As Figure 3 shows, this hypothesis is descriptively confirmed for all manipulations except responsibility and justice, as the differences in responsibility appear to be of equal magnitude in both scale scores, while the differences in justice appear to be greater with regard to the caring dimension. To test these hypotheses statistically, default Bayes factors (BFs) for repeated measurement analysis of variance designs (Rouder et al., 2012) were computed, comparing models with the main effects of the manipulation and dimensions with models containing these main effects and an additional interaction effect (with a *greater* difference for the manipulated dimension). The BF_{10} of these model comparisons exceeded 100 for all manipulations except for responsibility, indicating that the data at hand are much—indeed, over 100 times—more likely under the assumption of a model with interactions

(Etz & Vandekerckhove, 2018), which is usually judged to be “extreme evidence” (Lee & Wagenmakers, 2014). The BF_{10} for the responsibility text vignette equaled 1/7, which can be interpreted as some evidence for the model without interaction.

Intermediate Discussion of Study 1

The results of Study 1 initially showed a deviation of the five value dimensions derived from the literature. Thus, the manifest items (adjective pairs) did not sufficiently load on the theoretically proposed factor tolerance, so that only four dimensions can be represented empirically with 18 items in the resulting instrument (TIVO). Furthermore, the empirical evidence regarding the model fit pointed toward a second-order structure. This second-order factor (fairness) loads on the first-order factors justice, responsibility, and truthfulness. This can also be explained theoretically by the close relation of these constructs (Höffe, 2008).

Evidence from the second part (construct validation) of Study 1 showed mostly good construct validity, except for the justice and responsibility dimensions. Here, the manipulation of justice resulted in a greater difference in the caring dimension, and responsibility showed equal magnitude in both scale scores (fairness and caring). This leads to the question of whether the results of this manipulation are attributable to the instrument itself or to the text vignettes. Study 2 helps answer this question. There, the somewhat weak-structured procedure for item selection in Study 1 is compensated for by a strictly confirmatory and preregistered approach, which is presented in the next section.

Study 2 (Confirmatory Study)

After exploring the factor structure and construct validity of the new instrument in Study 1, we undertook Study 2 to gather more evidence for the factorial and construct validity of the TIVO. As the reliability of such results is generally

TABLE 2
Design of the Experiment in Study 2

Condition	Text vignettes describing fictitious teachers			
	1st vignette	2nd vignette	3rd vignette	4th vignette
Condition 1	ca_l	tr_h	re_h	ju_l
Condition 2	tr_h	re_l	ju_l	ca_h
Condition 3	re_l	ju_h	ca_h	tr_l
Condition 4	ju_h	ca_l	tr_l	r_h
Condition 5	ca_h	tr_l	re_l	ju_h
Condition 6	tr_l	re_h	ju_h	ca_l
Condition 7	re_h	ju_l	ca_l	tr_h
Condition 8	ju_l	ca_h	tr_h	re_l

Note. ca = caring; ju = justice; re = responsibility; tr = truthfulness; l = low specification; h = high specification.

threatened by several potential biases (Munafò et al., 2017), we planned Study 2 to have a strictly confirmatory nature and stated clearly defined research questions, a sampling rationale, and an analysis plan *before* assessing data (a process known as preregistration; Nosek et al., 2015); these elements were published along with the data on the Open Science Framework (<https://osf.io/bqnw9>).

Design

As the preregistration shows, Study 2 has two parts. The first is designed as a single-shot study assessing self-descriptions of values and aiming to replicate the factor structure from Study 1 using CFA. Part 2 is very similar to the analogous part of Study 1. In an incomplete rotated design, study participants were confronted with text vignettes describing fictitious teacher behavior with high or low manifestations of caring, justice, responsibility, or truthfulness, as per the design presented in Table 2.

Sample

The Study 2 sample was recruited through advertising in an obligatory lecture on educational science at a large German university. It consists of $N = 239$ student teachers (159 female, $M_{\text{semester}} = 1.28$, $SD_{\text{semester}} = 0.62$), whose participation was voluntary and unrewarded. The survey was carried out using paper-and-pencil procedures administered by trained test conductors as a groupwise assessment during coursework.

Materials

As in Study 1, the questionnaire consisted of two parts. The materials were adjusted with respect to the results of Study 1. In the first part, the participants were asked to describe themselves using the 18 adjective pairs (TIVO). In the second part, they judged the values of three teachers

described in text vignettes. We adapted the vignettes from Study 1 according to the results. The vignette describing teachers with high or low tolerance were omitted, and the vignette manipulating justice was redesigned. In each case, participants were asked to carefully read the vignette and to rate the described teacher using the 18 adjective pairs.

Results

To test the factor structure explored in Study 1 (see Table 3), we ran a series of CFA models with congeneric measurement models. We started with a model with only one factor as a reference (M2.1); we compared these results with a model with two factors based on our EFA results from Study 1 (M2.2). As Table 3 shows, the model implied covariance matrices from M2.2 that were much more similar to the empirical one for Model M2.2; this model was also preferred by the chi-square difference test. A model with the hypothesized second-order factor structure (M2.3; see Figure 2) again outperformed M2.2 concerning fit indices, which had not been the case for the final comparison of M2.3 with a model representing a four-factor structure (M2.4; see Figure 2). An analysis of internal consistencies led to very good results for three first-order dimensions and the second-order factor (caring $\omega = .826$; responsibility $\omega = .690$; justice $\omega = .776$; fairness $\omega = .802$) but to weak internal consistency for the first-order dimension of truthfulness ($\omega = .510$).

Manipulating the four first-order dimensions using text vignettes resulted in mean scores and standard deviations depicted in Figure 4 (see RDA for detailed tables). The mean patterns in this figure descriptively confirm our hypotheses. Manipulating the characteristics of the teachers in the vignettes concerning a specific dimension of the TIVO resulted in stronger mean differences of the corresponding second-order dimension in every vignette. Again, we computed BFs to test whether models including interaction terms predict the data better than models with only the main effects

TABLE 3
Results of the CFA in Study 2

Model	Fit indices					Comparison with model above	
	$\chi^2(df)$	CFI	TLI	RMSEA	SRMR	$\Delta\chi^2(df)$	<i>p</i>
M2.1	639.460 (129)	.581	.503	.140	.169	—	—
M2.2	275.259 (128)	.879	.856	.075	.082	364.201 (1)	<.001
M2.2	230.65 (126)	.914	.896	.064	.078	44.607 (2)	<.001
M2.3	227.994 (123)	.914	.893	.065	.077	2.659 (3)	.447

Note. CFA = confirmatory factor analyses; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

Means \pm 1SD of Caring and Fairness

Subplots indicate manipulated dimensions

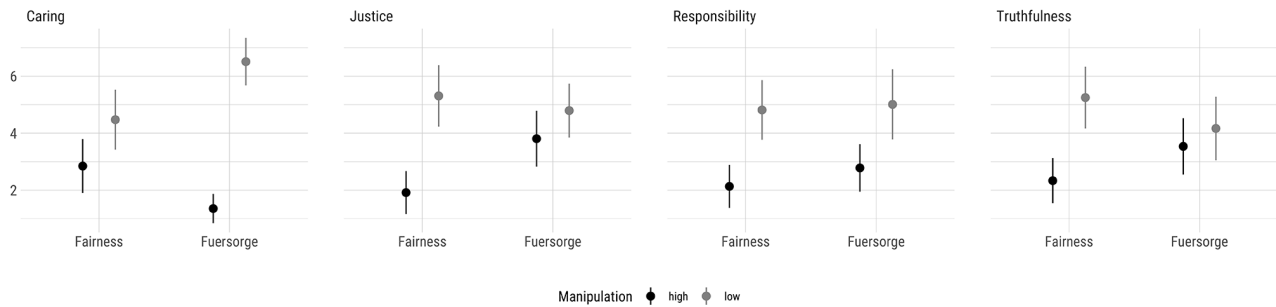


FIGURE 4. Effects of the manipulation (Study 2).

Note. Means and standard deviations were computed from mean scores of the fairness and caring scales.

of the dimension and manipulation. This resulted in “extreme evidence” for three vignettes ($BF_{10} > 100$) and moderate evidence for the manipulation of responsibility ($BF_{10} = 5.80$).

Intermediate Discussion of Study 2

The results from Study 2 confirmed the factor structure of the TIVO explored in Study 1. Teachers’ values can be categorized into the four dimensions caring, justice, responsibility, and truthfulness. The scores of three of these dimensions (justice, responsibility, and truthfulness) can be cumulated into a second-order factor called fairness. As Study 2 was preregistered, it can be interpreted as purely confirmatory. This again implies a high construct validity for the TIVO, as there is strong evidence for the appropriateness of the theoretically proposed dimensions (the CFA results) and adequate interpretations of the TIVO scores (the manipulated vignette results).

However, it should be noted that Study 2 was also based on a sample of only preservice teachers and that, while the samples of Studies 1 and 2 were independent of each other, the two studies were conducted with students from a single university.

Study 3 (Representative Study With In-Service Teachers)

One of the major limitations of Study 2 is the sample, which consists solely of preservice teachers from one university. To overcome this limitation, Study 3 seeks to confirm the factor structure of the TIVO using a representative sample of in-service teachers.

Design, Sample, and Materials

The TIVO as it resulted from Study 1 was used in an online survey conducted as part of a study with in-service teachers in the German states of North Rhine-Westphalia and Baden-Württemberg. A total of 254 (169 female) teachers in North Rhine-Westphalia and 154 (99 female) teachers in Baden-Württemberg were surveyed. Fifty of them were younger than 35, 94 between 35 and 44 years of age, 135 in the interval 45 to 54 years, and 160 older than 55 years. The average teaching experience was distributed as follows: 32 less than 5 years, 74 between 5 and 9 years, 157 between 10 and 19 years, 100 between 20 and 29 years, 74 with 30 years or longer experience. The survey was conducted in cooperation with a field service provider that routinely conducts

TABLE 4
Results of the CFA in Study 3

Model	Fit indices					Comparison with model above	
	$\chi^2(df)$	CFI	TLI	RMSEA	SRMR	$\Delta\chi^2(df)$	<i>p</i>
M3.1	903.049 (133)	.795	.795	.764	.092	—	—
M3.2	481.186 (132)	.907	.892	.078	.054	873.325 (1)	<.001
M3.3	428.957 (130)	.920	.906	.072	.051	36.120 (2)	<.001
M3.4	426.200 (127)	.920	.904	.073	.050	1.806 (3)	.614

Note. CFA incorporated sampling weights. CFA = confirmatory factor analyses; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

multitopic telephone surveys; the provider collected a sample of teachers using random-digit dialing and then asked respondents about participating in an online survey. Due to the random sampling, the distributions of age, sex, and type of school in our sample were very similar to official population statistics, which is why we deem our sample representative.

Results

In Study 3, we again used CFA to investigate the extent to which the factorial structure shown in Figure 2 is supplied by the data. However, as Study 3 relies on data incorporating sampling weights, we had to use appropriate methods to obtain correct estimates (Bollen et al., 2013). Therefore, we used functions of the lavaan.survey package (Oberski, 2014) and used pseudo-maximum likelihood for point and Taylor linearization for variance estimations. To test whether the proposed second-order structure of the TIVO is also supported by the representative data in Study 3, we fitted the same series of CFA models as used in Study 2. The results (see Table 4) provide strong evidence for the hypothesized second-order structure, as this model shows the best fit indices and, furthermore, is preferred based on likelihood ratio tests. A subsequent analysis of internal consistencies again shows very good results for two dimensions and the second-order factor (caring $\omega = .884$; responsibility $\omega = .768$; fairness $\omega = .955$) but only acceptable results for two first-order dimensions (truthfulness $\omega = .614$; justice $\omega = .602$).

Intermediate Discussion of Study 3

The strength of Study 3 consists in the representative data of in-service teachers. CFA based on this data again strengthens the construct validity of the TIVO. However, the data are based on only a pair of federal states in Germany, so a representative data set of all 16 federal states should be used for future analyses to reexamine the factor structure. Even if the analyses again show very good results for the two dimensions caring and responsibility and for the second-order factor of fairness, the reliability for the dimensions truthfulness and justice remains only acceptable.

Discussion

Values are considered important guidelines for thinking, feeling, and acting among teachers and are thus a key component of professionalism in teaching. To understand whether and how values can be developed in teacher education programs or how teachers' values affect their classroom behavior, it is necessary to have adequate possibilities for empirical investigation. This requires a clear definition of the construct and a valid operationalization, especially when studies claim to show an empirical relationship between values and teachers' classroom behavior (e.g., their influence on the choice and use of pedagogical strategies).

The specific relevance of being able to collect data according to teachers' values is also given by the fact that those values are regarded as an important facet of their professional competence (Kunter et al., 2013). While researching teachers' beliefs has a long empirical tradition (Fives & Buehl, 2012; Skott, 2015), there are still insufficient empirical instruments to address teachers' professional values. The TIVO is one of the first empirical instruments that can capture a broad range of profession-specific values. It can be used efficiently in future empirical research such as large-scale assessments due to its sleek design and extremely short processing time. This enables the prospective use of the instrument in other studies on teacher professionalism that seek to uncover connections between specific values and relevant actions, thoughts, or feelings in the teaching profession. To investigate the construct validity of the TIVO, we conducted three empirical studies. Below, the results of these studies are summarized and discussed with respect to methodological issues.

In a first step, five frequently mentioned profession-specific values and corresponding adjective and antonym pairs were extracted from the literature and validated by experts. This step also reveals one major limitation of our approach: Due to the fact that the literature from which values and adjectives were extracted was subjectively chosen by the authors, an exhaustive assessment of teachers' profession-specific values cannot be expected from the TIVO. Nor do we claim that the chosen values are either the most typical or

most important for the profession. In this regard, Delphi or bibliometric follow-up studies might provide further insights.

Based on the results of exploratory Study 1, we proposed a second-order factor structure for the profession-specific values of teachers (see Figure 2). Four first-order factors emerged: caring, justice, responsibility, and truthfulness. The latter three additionally serve as indicators for the second-order factor of fairness. The dimension of tolerance, which was also extracted from the literature, could not be found empirically as a separate dimension. In the independent and purely confirmatory preregistered Study 2, the empirical covariance structure of the data again showed the highest similarity to a theoretical covariance structure implied by the second-order structure (in comparison with other reasonable factorizations). While the first two studies were conducted at a single university and with only preservice teachers in their first semesters of study, Study 3 was based on a representative sample of in-service teachers in two noncontiguous German states. This study again provided evidence for the proposed second-order factor structure.

Overall, we deem our approach of conducting three independent studies with a clear distinction between exploratory and confirmatory study purposes to be methodologically rigorous (Makel & Plucker, 2014). Hence, the preference of the second-order structure that repeatedly emerged is evidence of the factorial validity of the TIVO (Piedmont, 2014) and, as the experimentally manipulated text vignettes induced TIVO ratings with the expected patterns, we gauge the TIVO's construct validity to be high (Messick, 1995). This appraisal is corroborated by a secondary analysis of the data from Study 2 (Drahmann et al., 2019). This analysis focused on convergent/divergent validity correlating the TIVO scales with the Portrait Values Questionnaire (Schwartz, 2006) dimensions and was also preregistered.



A few limitations associated with the development and current status of the TIVO must be considered. First, all three studies used samples from Germany. The extent to which the models can capture values in other countries, with their differing teacher education and school systems, is a question for future research. However, since the profession-specific values were derived from the international discourse, it is certainly conceivable that the values of teachers in other countries could also be recorded using the TIVO. Second, the TIVO of course cannot provide any information about the relevance of values for the development of competencies in the teaching profession, such as within teacher education. Further research is needed that uses the values as independent variables, along with others, that can operationalize teacher professionalism in a more complex and appropriate framework, such as professional knowledge (Shulman, 1987), motivation (Watt et al., 2012), and self-regulation (Jerusalem & Schwarzer, 1992). In such a larger context, the

specific relevance of teachers' values for teacher professionalism can be investigated. Third, an instrument like the TIVO will never be able to answer vital normative questions, such as the ethical or moral points of reference that are relevant for teachers and teacher education in different societies and their respective teacher education systems.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by a grant from Verband Bildung und Erziehung e.V. Behrenstrasse 24, 10117 Berlin, Germany

ORCID iDs

Samuel Merk  <https://orcid.org/0000-0003-2594-5337>
Colin Cramer  <https://orcid.org/0000-0003-3720-9708>

References

- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften [Keyword: Professional competencies of teachers]. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520. <https://doi.org/10.1007/s11618-006-0165-2>
- Bollen, K. A., Tueller, S. J., & Oberski, D. (2013). Issues in the structural equation modeling of complex survey data. In *Proceedings of the World Statistics Congress 2013* (pp. 1235–1240). <http://2013.isiproceedings.org/Files/STS010-P1-S.pdf>
- Bond, M. H. (1988). Finding universal dimensions of individual variation in multicultural studies of values: The Rokeach and Chinese value surveys. *Journal of Personality and Social Psychology*, 55(6), 1009–1015. <https://doi.org/10.1037/0022-3514.55.6.1009>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.
- Carr, D. (2006). Professional and personal values and virtues in education and teaching. *Oxford Review of Education*, 32(2), 171–183. <https://doi.org/10.1080/03054980600645354>
- Carr, D. (2010). Personal and professional values in teaching. In T. Lovat, R. Toomey, & N. Clement (Eds.), *International research handbook on values education and student wellbeing* (pp. 63–74). Springer. <https://doi.org/10.1007/978-90-481-8675-4>
- Clayman, C. S. (1961). Values and the teacher. *Journal of Education*, 143(1), 23–27. <https://doi.org/10.1177/002205746114300403>
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. <https://doi.org/10.1037/h0040957>
- Drahmann, M., Merk, S., & Cramer, C. (2019). Werthaltungen im Lehrerberuf. Forschungsstand zu deren Erfassung und Konstruktvalidierung des “Tübingen Inventory for Measuring Value Orientation in the Teaching Profession” (TIVO) [Value orientations in the teaching profession. Current approaches in instrument development and convergent/divergent validation of the “Tübingen Inventory for Measuring Value Orientation in the Teaching Profession” (TIVO)]. In C. Rotter, C. Schülke, & C. Bressler (Eds.), *Lehrerhandeln—eine Frage der Haltung? [Teachers' conduct—A question of attitudes?]* (pp. 174–193). Beltz Juventa.

- Dunn, T. J., Baguley, T., & Brunsten, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412. <https://doi.org/10.1111/bjop.12046>
- Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J., & Midgley, C. (1983). Expectancies, values and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives* (pp. 75–146). Freeman.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*(1), 109–132. <https://doi.org/10.1146/annurev.psych.53.100901.135153>
- Etz, A., & Vandekerckhove, J. (2018). Introduction to Bayesian inference for psychology. *Psychonomic Bulletin & Review*, *25*, 5–34. <https://doi.org/10.3758/s13423-017-1262-3>
- Fives, H., & Buehl, M. M. (2012). Spring cleaning for the “messy” construct of teachers’ beliefs: What are they? Which have been examined? What can they tell us? In K. R. Harris, S. Graham, T. Urdan, S. Graham, J. M. Royer, & M. Zeidner (Eds.), *APA educational psychology handbook: Vol. 2. Individual differences and cultural and contextual factors* (pp. 471–499). American Psychological Association. <https://doi.org/10.1037/13274-019>
- Fries, S., Schmid, S., & Hofer, M. (2007). On the relationship between value orientation, valences, and academic achievement. *European Journal of Psychology of Education*, *22*(2), 201–216. <https://doi.org/10.1007/BF03173522>
- Gudmundsdottir, S. (1990). Values in pedagogical content knowledge. *Journal of Teacher Education*, *41*(3), 44–53. <https://doi.org/10.1177/002248719004100306>
- Harder, P. (2014). *Werthaltungen und Ethos von Lehrern. Empirische Studie zu Annahmen über den guten Lehrer* [Values and ethos of teachers. An empirical study of assumptions about the good teacher]. Otto-Friedrich-Universität Bamberg.
- Heim, E., Scholten, S., Maercker, A., Xiu, D., Cai, D., Gao, Z. H., Lu, S., Sang, Z. Q., Wei, J., Kochetkov, Y., & Margraf, J. (2017). Students’ value orientations in contemporary China: Analysis of measurement invariance and latent mean differences in comparison with students from Germany and Russia. *Journal of Cross-Cultural Psychology*, *48*(4), 511–531. <https://doi.org/10.1177/0022022117696800>
- Höffe, O. (2008). *Lexikon der Ethik* (7th ed.). C. H. Beck.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. <https://doi.org/10.1007/BF02289447>
- Horton, J. (1998). Toleration. In E. Craig (Ed.), *Routledge encyclopedia of philosophy: Vol. IX* (pp. 429–433). Routledge.
- Inglehart, R., Basanez, M., Diez-Medrano, J., Halman, L., & Luijckx, R. (2000). *World values surveys and European values surveys, 1981-1984, 1990-1993, and 1995-1997*. Institute for Social Research, University of Michigan.
- James, M., & McCormick, R. (2009). Teachers learning how to learn. *Teaching and Teacher Education*, *25*(7), 973–982. <https://doi.org/10.1016/j.tate.2009.02.023>
- Jerusalem, M., & Schwarzer, R. (1992). Self-efficacy as a resource factor in stress appraisal processes. In R. Schwarzer (Ed.), *Self-efficacy: Thought control of action* (pp. 195–213). Hemisphere.
- Klaassen, C. A., Osguthorpe, R. D., & Sanger, M. N. (2016). Teacher education as a moral endeavor. In J. Loughran, & M. Hamilton (Eds.), *International handbook of teacher education* (pp. 523–557). Springer. https://doi.org/10.1007/978-981-10-0366-0_14
- Kluckhohn, C. (1951). Values and value-orientations in the theory of action: An exploration in definition and classifications. In T. Parsons, & E. A. Shils (Eds.), *Toward a general theory of action* (pp. 388–433). Harvard University Press. <https://doi.org/10.4159/harvard.9780674863507.c8>
- Kohlberg, L. (1981). *The philosophy of moral development: Moral stages and the idea of justice*. Harper & Row.
- Krippendorff, K. (2019). *Content analysis. An introduction to its methodology* (4th ed.). https://doi.org/10.1111/j.1468-4446.2007.00153_10.x
- Krobath, H. T. (2009). *Werte: Ein Streifzug durch Philosophie und Wissenschaft* [Values: A digression through philosophy and science]. Königshausen und Neumann.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Eds.). (2013). *Cognitive activation in the mathematics classroom and professional competence of teachers. Results from the COACTIV project*. Springer. <https://doi.org/10.1007/978-1-4614-5149-5>
- Lauermaun, F., & Karabenick, S. A. (2013). The meaning and measure of teachers’ sense of responsibility for educational outcomes. *Teaching and Teacher Education*, *30*(February), 13–26. <https://doi.org/10.1016/j.tate.2012.10.001>
- Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139087759>
- Lindeman, M., & Verkasalo, M. (2005). Measuring values with the short Schwartz’s Value Survey. *Journal of Personality Assessment*, *85*(2), 170–178. https://doi.org/10.1207/s15327752jpa8502_09
- Mägdefrau, J. (2008). Welche Werte haben zukünftige Lehrer/innen? Lehramtsstudierende und Studierende nicht pädagogischer Fachrichtungen im Vergleich [Which values do teachers-to-be have? Student teachers and students of other disciplines in comparison]. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, *28*(1), 36–55.
- Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, *43*(6), 304–316. <https://doi.org/10.3102/0013189X14545513>
- Marini, M. M. (2000). Social norms and values. In E. F. Borgatta, & R. J. V. Montgomery (Eds.), *Encyclopedia of sociology* (2nd ed.; pp. 2828–2840). Macmillan.
- 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: A Multidisciplinary Journal*, *11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons’ responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A

- manifesto for reproducible science. *Nature Human Behaviour*, 1, Article 0021. <https://doi.org/10.1038/s41562-016-0021>
- Nagengast, B., Trautwein, U., Kelava, A., & Lüdtke, O. (2013). Synergistic effects of expectancy and value on homework engagement: The case for a within-person perspective. *Multivariate Behavioral Research*, 48(3), 428–460. <https://doi.org/10.1080/00273171.2013.775060>
- Nash, R. J. (1991). Three conceptions of ethics for teacher educators. *Journal of Teacher Education*, 42(3), 163–172. <https://doi.org/10.1177/002248719104200302>
- Noddings, N. (1984). *Caring: A feminine approach to ethics and moral education*. University of California Press.
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., . . . Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348(6242), 1422–1425. <https://doi.org/10.1126/science.aab2374>
- Oberski, D. (2014). lavaan.survey: An R package for complex survey analysis of structural equation models. *Journal of Statistical Software*, 57(1), 1–27. <https://doi.org/10.18637/jss.v057.i01>
- Oser, F. K. (1994). Chapter 2: Moral perspectives on teaching. *Review of Research in Education*, 20(1), 57–127. <https://doi.org/10.3102/0091732X020001057>
- Oser, F. K. (1998). *Ethos—die Vermenschlichung des Erfolgs. Zur Psychologie der Berufsmoral von Lehrpersonen* [Ethos—the humanization of success. On the psychology of professional morals of teachers]. VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-322-97398-6>
- Oser, F. K., & Biedermann, H. (2018). The professional ethos of teachers. Is only a procedural discourse approach a valid model? In A. Weinberger, H. Biedermann, J.-L. Patry, & S. Weyringer (Eds.), *Professionals' ethos and education for responsibility* (pp. 23–39). Brill. https://doi.org/10.1163/9789004367326_003
- Piedmont, R. L. (2014). Factorial validity. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 2148–2149). https://doi.org/10.1007/978-94-007-0753-5_984
- Revelle, W., & Rocklin, T. (1979). Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioral Research*, 14(4), 403–414. https://doi.org/10.1207/s15327906mbr1404_2
- Rokeach, M. (1973). *The nature of human values*. Free Press. <https://doi.org/10.1093/sw/19.6.758>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, 56(5), 356–374. <https://doi.org/10.1016/j.jmp.2012.08.001>
- Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology*, 25, 1–65. [https://doi.org/10.1016/S0065-2601\(08\)60281-6](https://doi.org/10.1016/S0065-2601(08)60281-6)
- Schwartz, S. H. (2006). A theory of cultural value orientations: Explication and applications. *Comparative Sociology*, 5(2–3), 137–182. <https://doi.org/10.1163/156913306778667357>
- Schwartz, S. H. (2007). Value orientations: Measurement, antecedents and consequences across nations. In R. Jowell, C. Roberts, R. Fitzgerald, & G. Eva (Eds.), *Measuring attitudes cross-nationally: Lessons from the European Social Survey* (pp. 169–203). Sage.
- Schwartz, S. H. (2011). Values: Cultural and individual. In F. J. R. van de Vijver, A. Chasiotis, & S. M. Breugelmans (Eds.), *Fundamental questions in cross-cultural psychology* (pp. 463–493). Cambridge University Press. <https://doi.org/10.1017/CBO9780511974090.019>
- Schwartz, S. H., & Cieciuch, J. (2021). Measuring the refined theory of individual values in 49 cultural groups: Psychometrics of the revised Portrait Value Questionnaire. *Assessment*. Advance online publication. <https://doi.org/10.1177/1073191121998760>
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, 54(2), 93–105. <https://doi.org/10.1037/0003-066X.54.2.93>
- Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–23. <https://doi.org/10.17763/haer.57.1.j463w79r56455411>
- Skott, J. (2015). The promises, problems, and prospects of research on teachers' beliefs. In H. Fives, & M. G. Gill (Eds.), *International handbook of research on teachers' beliefs* (pp. 13–30). Routledge.
- Thayer-Bacon, B. J. (2008). Caring reasoning. In D. Fasko, & W. Willis (Eds.), *Contemporary philosophical and psychological perspectives on moral development and education* (pp. 83–105). Hampton Press.
- Vernon, P. E., & Allport, G. W. (1931). A test for personal values. *Journal of Abnormal and Social Psychology*, 26(3), 231–248. <https://doi.org/10.1037/h0073233>
- Veugelers, W. (2010). Moral values in teacher education. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education* (3rd ed.; pp. 650–655). Elsevier. <https://doi.org/10.1016/B978-0-08-044894-7.00635-7>
- Wannamaker, M., & Tennyson, W. W. (1970). The value orientation of beginning elementary teacher education students. *Journal of Teacher Education*, 21(4), 544–550. <https://doi.org/10.1177/002248717002100416>
- Watt, H. M. G., Richardson, P. W., Klusmann, U., Kunter, M., Beyer, B., Trautwein, U., & Baumert, J. (2012). Motivations for choosing teaching as a career: An international comparison using the FIT-Choice scale. *Teaching and Teacher Education*, 28(6), 791–805. <https://doi.org/10.1016/j.tate.2012.03.003>
- Weinberger, A., Biedermann, H., Patry, J.-L., & Weyringer, S. (Eds.). (2018). *Professionals' ethos and education for responsibility*. Brill. <https://doi.org/10.1163/9789004367326>
- Willis, G. B. (2005). *Cognitive interviewing. A tool for improving questionnaire design*. Sage. <https://doi.org/10.4135/9781412983655>
- Wray-Lake, L., Christens, B. D., & Flanagan, C. A. (2014). Community values. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 1102–1107). Springer. https://doi.org/10.1007/978-94-007-0753-5_482
- Zhu, W., & Chen, A. (2018). Value orientation inventory: Development, applications, and contributions. *Kinesiology Review*, 7(3), 206–210. <https://doi.org/10.1123/kr.2018-0030>
- Zia, R. (2007). Values, ethics and teacher education: A perspective from Pakistan. *Higher Education Management*

and Policy, 19(3), 105–125. <https://doi.org/10.1787/hemp-v19-art20-en>

Authors

SAMUEL MERK is an educational researcher at the University of Education Karlsruhe. His main research interests lie in teacher beliefs and values and teachers dealing with evidence.

MARTIN DRAHMANN was an educational researcher at the University of Tübingen. He died completely unexpectedly in 2019.

COLIN CRAMER is an educational researcher at the University Tübingen. He aims to understand who teachers and school leaders are, under which circumstances they work, and how they can be prepared for their tasks appropriately.