The Teacher Self-Efficacy Scale (TSES) factorial structure evidence review and new evidence from Polish speaking samples

Abstract

The Teacher Self-Efficacy Scale (TSES) is a widely accepted and enthusiastically used measurement instrument. Unfortunately, little firm evidence is available to support its factorial structure. The article reports on the results of the TSES administered to large representative samples of primary (2,083) and lower-secondary school teachers (2,382) in Poland, along with the results of reanalyzed U.S. data, based on which the scale was developed. The results support TSES as a reliable instrument applicable across countries. However, they also highlight the importance of analyzing the factorial structure of the scale with each application, as it may require modifications when administered to different samples. Additionally, moderate evidence was found in Polish data for TSES to be invariant between primary and lower-secondary school teachers.

Keywords: Teacher Self-Efficacy Scale (TSES), self-efficacy, teachers, factorial structure, measurement invariance

Introduction

Various studies demonstrate that of the factors controlled by the school system teachers have the strongest impact on student achievement (Akiba, LeTendre, & Scribner, 2007; Rivkin, Hanushek, & Kain, 2005; Taylor, Roehrig, Hensler, Connor, & Schatschneider, 2010). It is not clear, however, which teacher attributes or behaviors in particular explain the differences in teaching effectiveness. Since teacher self-efficacy is empirically linked with student
outcomes (Armor et al., 1976), it has become extensively studied as an explanatory variable of differences in teaching effectiveness.

Teacher self-efficacy is a psychological construct, defined as the teachers’ confidence in their ability to enhance students’ motivation and boost learning outcomes (Bandura, 1977). The origins of the concept can be found in Rotter’s (1966) locus of control theory. Teacher self-efficacy can be also viewed as conceptually and empirically overlapping with self-esteem and emotional stability (Bono & Judge, 2003). In this paper, the focus is on individual teacher self-efficacy in the practice of teaching in general as opposed to collective teacher efficacy as explored in much of the literature (Gibbs & Powell, 2012; Goddard, Hoy, & Woolfolk Hoy, 2004, Moolenaar, Sleegers, & Daly, 2012) and efficacy in teaching particular subjects, e.g. maths (Holzberger, Philipp, & Kunter, 2013), physical education (Pan, 2013), science (Kılınç et al., 2013) or music (Garvis, 2013).

Teacher self-efficacy has been considered not only as having a direct effect on students’ outcomes (Caprara, Barbaranelli, Steca, & Malone, 2006; Mohamadi & Asadzadeh, 2012), but also as influential indirectly by increasing students’ motivation to learn (Midgley, Feldlaufer, & Eccles, 1989) and raising their self-efficacy, that is confidence in their own ability to do well at school (Anderson, 1988). Teacher self-efficacy has been shown to correlate positively with teachers’ engagement and job satisfaction (Caprara, Barbaranelli, Borgogni, & Steca, 2003; Skaalvik & Skaalvik, 2014), and negatively with job burnout (Aloe, Amo, & Shanahan, 2014; Yu, Wang, Zhai, Dai, & Yang, 2015). Self-efficacy has also been found to be a strong predictor of teachers’ psychological well-being, physical health, and quitting intentions (Wang, Hall, & Rahimi, 2015). Teachers with a higher sense of self-efficacy show more support and provide a more positive classroom environment than teachers with lower self-efficacy (Guo, Connor, Yang, Roehrig, & Morrison, 2012). They also engage more in relationships with other teachers and parents (Coladarci, 1992; Hoover-Dempsey,
Zee and Koomen (2016), through their synthesis of 40 years of teacher self-efficacy research, showed more consequences that teacher self-efficacy has on a broad range of outcomes.

Serious problems with measuring teacher self-efficacy have been encountered since the concept first appeared in scholarly discourse. Tschannen-Moran, Woolfolk Hoy and Hoy (1998) surveyed studies on teacher efficacy, conducted from 1974 to 1997. Most of the instruments reviewed turned out to be unacceptable to the community of researchers or unpopular due to an analytically troublesome question format. There were also concerns about their low reliability. On the other hand, measures which attracted researchers’ attention, such as the Teacher Efficacy Scale (TES) by Gibson and Dembo (1984), generated several statistical and interpretational problems (Tschannen-Moran & Woolfolk Hoy, 2001, p. 789).

Klassen, Tze, Betts and Gordon, (2011) took up Tschannen-Moran, Woolfolk Hoy and Hoy (1998) in their investigation of the state of teacher efficacy research from 1998 to 2009. Despite the warnings offered by Henson (2002), Henson, Kogan and Vacha-Haase (2001), as well as Tschannen-Moran and Woolfolk Hoy (2001), the conceptually troubled scale authored by Gibson and Dembo (1984) was used in almost one third of teacher self-efficacy studies reviewed by Klassen et al. (2011). Other instruments were flagged as conceptually incongruent with a theory-based understanding of the concept (Betoret, 2006; Evers, Tomic, & Brouwers, 2005; Friedman & Kass, 2002; Somech & Drach-Zahavy, 2000; Tobin, Muller, & Turner, 2006; Tournaki & Podell, 2005). The main conceptual difficulty of the abovementioned measures is a focus on the teachers’ beliefs about their control of students’ outcomes (originating in locus of control theories) rather than a focus on the teachers’ capabilities to effectively teach students (Klassen et al., 2011, p. 36). In the publications the authors reviewed, they found that only a few measures showed close congruence with Bandurian theory (Caprara, Barbaranelli, Borgogni, Petitta, et al., 2003; Caprara,
Barbaranelli, Borgoni, & Steca, 2003; Tschannen-Moran & Woolfolk Hoy, 2001). Authors concluded: “Much of the teacher efficacy research continues to use discredited, poorly conceptualized and flawed measures. (...) Findings from the studies using flawed measures can lead to misleading conclusions, as well as a kind of definitional entropy where the meaning of carefully defined psychological constructs lose precision over time, eventually losing predictive power and theoretical distinctiveness” (Klassen et. al, 2011, p. 37). Said this, they follow with recommendation of the TSES over other scales. It was found to be “considerably more congruent with self-efficacy theory than many of the other measures” (Klassen et al., 2011, p. 40).

The purpose of this article is to summarize the evidence regarding the factorial structure of the TSES from both English and non-English speaking samples. Secondly, the psychometric characteristics of TSES are reported from the original Tschannen-Moran and Woolfolk Hoy (2001) data on 255 in-service teachers, and Polish data on 2,083 primary and 2,382 lower-secondary schools teachers.

Review of the evidence for the TSES factorial structure

Despite the fact that the TSES is widely used, sadly there is very little evidence on its factorial structure and psychometric characteristics. According to a search of the Web of Science Core Collection, in the period between 2001 and 2014 the article by Tschannen-Moran and Woolfolk Hoy (2001) introducing the TSES was cited in 356 reviewed articles and proceedings papers. Only a few empirical works investigated the scale factorial structure or at least assessed its reliability on a studied sample. Of interest to us are only publications in which the results of a confirmatory factor analysis (CFA) are reported. CFA investigates whether the empirical data fit to a specified theoretical model; in this case, whether the data fit the three-factor structure, identified by Tschannen-Moran and Woolfolk Hoy (2001). Despite
the authors originally assuming three factors as orthogonal, what I refer to here as an “original three-factor structure, identified by the TSES originators”, it is a simple confirmatory oblique three-factor model, henceforth a baseline model. Firm evidence comes from 9 applications of the TSES English version. Another 6 evidence-supported studies with CFA results from non-English speaking samples are available. The evidence is summarized in Table 1.

Model fit was assessed with the most commonly used (McDonald & Ho, 2002) fit indices, i.e. root mean square error of approximation (RMSEA) and comparative fit index (CFI). Additionally, the standardized root mean square residuals (SRMR) are reported, since they are also recommended (e.g., Boomsma, 2000; Hu & Bentler, 1999; Kline, 2005; Mueller & Hancock, 2008). After Hu and Bentler (1999) I assumed that RMSEA values not higher than .060, CFI values not lower than .950, and SRMR values not higher than .080 indicate good fit. The confidence intervals were computed for RMSEA when results of my analysis are reported. Ideally a lower value of the confidence interval is close to zero and an upper value does not exceed .080.

The TSES comprises three subscales, representing the teacher self-efficacy dimensions: instructional strategies self-efficacy (ISSE), classroom management self-efficacy (CMSE), and student engagement self-efficacy (SESE). The scale is available in short (4 items per subscale) and long forms (8 items per subscale). For details of which items are used in short form please see Table 5 in Appendix.

|Table 1|

Three research questions were posed before reviewing in detail the selected works: (1) whether the proposed by TSES originators three-factor structure holds; (2) whether the model
requires any modifications to fit data; (3) whether the factors are highly correlated, which
would suggest the presence of a general factor representing a general teacher efficacy, or
should they be viewed as orthogonal?

The high reliability of the subscales, comprising the TSES long form, was reported in the
reviewed works. For a ISSE mean $\alpha$ equals .90, and for CMSE .89, SESE .87. The $\chi^2$/df ratio
was on average 3.22, where values below 3 indicate a good fit. Please note that none of the
reviewed studies reported the robust $\chi^2$. The RMSEA for the baseline model was ranging from
.053 to .134 with its average value of .077, slightly exceeding the upper boundary of
acceptability. The average CFI of .937, ranging from .820 to .990, almost reached the
threshold recommended by Hu and Bentler (1999). To sum up, evidence is available to
support the original three-factor structure of the TSES (studies no. 2, 3, 4, 5, 7-8). However,
some studies failed to replicate it (6, 9, 11, 12, 15). Please note that eight studies (4, 5, 10-15)
investigated short form of the scale.

The second question posed was whether the baseline model required any amendments to fit
the investigated sample. In several studies, the baseline model did not replicate well, and
therefore required rather drastic modifications. Wolters and Daugherty (2007) were able, by
dropping six items and allowing error covariances between four pairs of items, to
considerably improve model fit from $\chi^2(249) = 1400.62$; RMSEA = .100; CFI = .860 to
$\chi^2(128) = 577.65$, RMSEA = .080; CFI = .930. In another study, Bosma, Hessels and Resing
(2012) adapted the model by eliminating three items which loaded on more than one factor
and allowing error covariances between three pairs of items, because of an assumed content
overlap of items. The robust method showed a more reasonable fit (SBy$\chi^2(183) = 237.75$;
RMSEA = .038; CFI = .950) compared to the baseline model ($\chi^2(249) = 551.23$; RMSEA = .070;
CFI = .820). In the Klassen et al. (2009) study, it was only the model for the U.S.
sample that did not require any changes. The fit statistics improved considerably in non-U.S.
samples by allowing error covariances between the last two items of the scale short form. As
argued by the authors, the correlated errors may be a function of translation, whereby some
items seem closer in meaning in some languages than in the original version or a function of
proximity (items placement at the end of the scale). Other possible explanations of the
correlated errors of items placed at the end of the scale are respondents’ fatigue (respondents
were more tired while answering last items) or speediness (respondents speed up towards the
end of the scale, answering carelessly).

The third question asked was whether the factors are highly correlated, thus researchers
should lean more toward models with general factor presence. In all the reviewed works (see
Table 1) factors correlate typically above .6, which suggests that they should not be
considered as orthogonal. Therefore, it can be concluded after Brown (2006) that due to
strong inter-factor correlations, it is wise to treat the TSES as a measure of teachers’ general
sense of self-efficacy.

Lastly, it is worth mentioning the limitation of the studies reviewed, i.e. the low quality of the
samples used. With the exception of the Heneman, Kimball and Milanowski (2006) study, all
the analyses included in this review were based on the data from rather small convenience
samples. Additionally, the majority of the analyses were run on mixed-teacher samples. There
are no results available of analyses that were run on separate samples of primary and lower-
secondary school teachers in one country, allowing for scale invariance test between these
two groups. The present study hopes to fill in this gap.

Materials and Methods

Polish adaptation of the TSES
The TSES Polish adaptation was performed as follows. First, the scale was translated by three psychologists (PhDs, who familiarized themselves with articles on TSES adaptation in other countries) and two native translators (a pedagogue and an English philologist, who did not read such articles). Each person independently produced a proposal for the translated scale along with a short report justifying the decisions made and doubts remaining. Second, the translators met in joint session during which the common version of the translation was developed consensually. Third, the scale was back-translated independently by two translators who did not know the scale in the original English version. Fourth, all materials were submitted to three different experts (pedagogy, English philologist, test adaptation specialist). During a series of meetings experts developed the version of the scale, further tested in a pilot study. The scale was piloted in a three-stage study: first among 10 teachers and 3 bilingual persons, second on 118 teachers, third on 569 teachers.

**Data sources and sampling**

Tschannen-Moran and Woolfolk Hoy (2001) reported the TSES characteristics based on a combined sample of pre-service and in-service teachers (410 teachers, including 103 pre-service teachers, 255 in-service teachers, and 38 respondents who failed to indicate their teaching experience). The authors found that the responses of pre-service teachers were unstable and that their sense of their own abilities was not sufficiently well-formed for the scale structure to hold. Therefore, in this article the TSES was reanalyzed, using the original Tschannen-Moran and Woolfolk Hoy (2001) data on in-service teachers only.

This study also uses data from longitudinal studies: the Educational Value-Added for Primary Schools (EVA-PS) study, run between 2009 and 2015, and Lower-Secondary Schools (EVA-LSS), run between 2010 and 2012 in Poland. Both studies focused on student achievement predictors, including teacher effect; thus students were sampled first and the teachers’ samples...
depended on the students’ samples. The sampling procedure was similar in both studies. Stratified two-stage cluster sampling was performed on students. First, public schools (primary and lower-secondary schools separately) were split into strata based on their location (village, town, city) and the number of Grade 1 classes in a school. Special education schools and schools with fewer than ten (in the case of primary schools) or twenty (in the case of lower-secondary schools) students were excluded from the sampling frame. Within strata, schools were selected randomly with a probability proportional to the number of students in Grade 1 classes. Second, two Grade 1 classes were sampled from each school (if a school ran only one or two Grade 1 classes, all classes were included into the sample). It should be noted that a class is a stable group of students taught by different teachers for each subject. All teachers in the sampled schools were invited to the study, therefore school sampling weights were assigned to the teachers. The results presented in this article are representative of the Polish primary and lower-secondary school teacher populations.

Data collection

This article reports on the results of two applications of the TSES Polish version. The first time the TSES was surveyed among primary school teachers (May-June 2012). Data were collected on 2,295 teachers, teaching 4,942 students in 180 schools. The response rate reached

---

1 I argue that the teachers’ samples representativeness is a given, since it follows students’ random samples. Proofs for the representativeness of the student samples can be found in Dolata et al. (2013) for EVA-PS and in Dolata et al. (2015) for EVA-LSS. Very limited data on the teacher population, which can serve as the criterion for a representativeness-check of the teacher samples, is provided by the Central Statistical Office (2012). In the population of primary school teachers 86% were females, 3% were trainees, 16% contractual, 26% appointed, and 55% were chartered teachers in the school year 2011/12, while in the sample the figures were 94%, 2%, 8%, 21%, and 64%, respectively (remaining 5% are missing data). In the population of lower-secondary school teachers 75% were females, 3% were trainees, 17% contractual, 26% appointed, and 54% were chartered teachers, while in the sample 88%, 3%, 15%, 23%, and 55%, respectively (remaining 4% are missing data). The differences between the samples and the populations do not undermine the samples’ representativeness, since numbers provided by the Central Statistical Office are expressed in terms of full-time employees, while characteristics of the samples are expressed in terms of teachers, regardless contract type.
90%. The final analyses were performed on the 2,083 teachers after the exclusion of 9% observations due to missing data on all 24 TSES items, null variance across all 24 items, or missing information on the school ID variable. The second application of the TSES was performed on lower-secondary school teachers (April-June 2012). Data were collected on 2,630 teachers, teaching 5,567 students in 150 schools. The response rate reached 89%. For the final analyses data on 2,382 teachers was retained, also excluding 9% of observations due to the abovementioned reasons.

Pen-and-paper questionnaires with the Polish version of the TSES long form (see Table 5 in the Appendix) were used. Responses were provided on a 9-point Likert-type scale. The whole range of the scale was used by respondents for each item. However, the distributions of answers were left skewed, with a cumulative percentage of the indications of the first three categories ranging from 0.24% to 1.83% depending on sample. This may suggest shortening the scale and using more adequate analytic approaches such as categorical CFA (CCFA) or the graded response model (GRM). Nevertheless, since in none of the reviewed studies were the TSES items analysed as categorical data, for comparability in this paper responses were also treated as continuously distributed.

**Sample characteristics**

The average age of primary school teachers in the Polish sample was 45, of lower-secondary school teachers 42, and of teachers in the U.S. sample 35. The average teaching experience was 21, 16, and 8 years respectively. Females comprised 94% of the Polish primary school teacher sample, 88% of the lower secondary school teacher sample, and 67% of the U.S. sample. In the U.S. sample, 14 teachers were in preschool (teaching 3-5 years old children), 80 in elementary schools (6-10), 70 in middle schools/Jr. high (11-13) and 75 in high schools (14-17). There were 2,083 primary school (7-12) teachers and 2,382 lower-secondary school
(13-15) teachers in Polish samples. In the U.S. sample, 29% were maths teachers, 16% science teachers, 18% language teachers, and 5% foreign language teachers. In the Polish primary school teacher sample, the proportions were 16%, 12%, 18%, and 2%, respectively, while in the lower-secondary school teacher sample they were 17%, 28%, 19%, and 27%, respectively. In the Polish primary school teacher sample 53% were teachers of Grade 1-3 students (7-9), teaching all subjects.

Results

Reanalysis of the original Tschannen-Moran and Woolfolk Hoy (2001) data

For the presentation of the scale’s psychometric characteristics, Tschannen-Moran and Woolfolk Hoy (2001) used the results of principal-axis factoring, with associated eigen values greater than one, followed by varimax rotation. The problems with such an approach were described by its originator (Kaiser, 1970) and illustrated by example (Armstrong, 1967). Preacher and MacCallum (2003) aptly pointed out that when carrying out exploratory factor analysis (EFA) it is wise to assume that factors are correlated, rather than force an orthogonal solution. If an optimal simple structure is exhibited by orthogonal factors, an obliquely rotated factor solution will resemble an orthogonal one anyway (Floyd & Widaman, 1995).

Following these suggestions I used the direct oblimin rotation method for oblique rotation (Jennrich & Sampson, 1966) of the factor structure proposed by Tschannen-Moran and Woolfolk Hoy (2001). The analysis revealed that the assumption of uncorrelated factors proved to be incorrect. Factors’ correlations were strong, ranging from .47 to .71.

The scale and item-level characteristics presented by Tschannen-Moran and Woolfolk Hoy (2001) in their article are lacking in the evidence from the CFA. The CFA model was run on the original Tschannen-Moran and Woolfolk Hoy (2001) data on in-service teachers. The
model (as well as other models presented here) was estimated in Mplus 8.0 (Muthén & Muthén, 1998-2017) by maximum likelihood estimation with robust (Huber, 1967; White, 1980) standard errors (MLR). Out of 24 items, only five were normally distributed. Based on the Doornik-Hansen (2008) test, Mardia’s (1970) test and the Henze-Zirkler (1990) test, the null hypothesis of multivariate normality was rejected for the remaining items. Since the empirical item distributions were non-normal, inflated ML $\chi^2$ goodness-of-fit test statistics and underestimated parameter standard errors were expected (Muthén & Kaplan, 1985; Kaplan, 2000). The use of robust standard errors estimates and the Satorra and Bentler (2010) scaled $\chi^2$ (SB$\chi^2$) helped rule out these biases in assessment of a model’s accuracy. Because the teachers who participated in the Tschannen-Moran and Woolfolk Hoy (2001) study had not been selected randomly, no sampling weights were available.

A baseline model (representing factorial structure proposed by Tschannen-Moran & Woolfolk Hoy, 2001) were found to fit the data acceptably based on SRMR but not RMSEA and CFI values (RMSEA (90%CI) = .069 (.062–.077); CFI = .871; SRMR = .064; for details see Mod.1 in electronic supplementary materials, ESM). The management factor correlates with the instructional factor at the level of .682, engagement with the instructional factor .645 and with the management factor .596. Factor loadings and their standard errors are presented in Table 2. Running the second-order factor model allowed for the assessment of loadings of the specific factors with the general second order factor. The instructional, management and engagement factors loaded the second-order general factor by .859, .793, .751, respectively. High between-factor correlations, as well as the emergence of the second-order factor, indicate that the TSES measures the underlying general construct of self-efficacy.

<Table 2>
Factorial structure of the TSES in the Polish speaking samples

The results presented above have not taken into account the clustering of teachers within schools. While the original data of Tschannen-Moran and Woolfolk Hoy (2001) on in-service teachers does not allow us to account for clustering (lack of school ID variable), the Polish data does. All results from Polish samples presented in this article are not only robust to the non-normality (by using the MLR estimator), but also to the non-independence of teachers clustered within schools (by adjusting the standard errors using a sandwich estimator; Muthén & Muthén, 1998-2017; Muthén & Satorra, 1995).

Based on RMSEA and SRMR but not CFI values, the baseline model fits satisfactorily both primary (RMSEA (90%CI) = .055 (.052–.057); CFI = .881; SRMR = .052; see Mod.2 in ESM) and lower-secondary school teacher data (RMSEA (90%CI) = .059 (.057–.061); CFI = .885; SRMR = .058; see Mod.3 in ESM). The results of EFA ran on Polish data indicated that item 12 was loading strongly not only on the classroom management factor but also the instructional factor. This was most likely due to a shift in the translation of item 12 towards instructional strategies. Items 20, 21, 23 were detected as having strong loadings on the instructional factor and on the student engagement factor. Hence, items 12 and 23 were set to load the instructional factor, too. Doing the same for items 20 and 21 failed to improve the model fit considerably. Additionally, based on high modification indices, error covariances between items 6 and 12, 3 and 23, 1 and 6 were allowed. Allowing error covariances can be justified here, because in the pen-and-paper questionnaire the correlated items were placed in pairs one by one. If a computer-based questionnaire were to be administered with random item ordering, no error covariances within these pairs should be expected. This is, however, a hypothesis to be checked in further studies. The model, according to the specification described above, proved to fit considerably better the primary school teacher data (RMSEA (90%CI) = .044 (.042–.047); CFI = .924; SRMR = .043; see Mod.4 in ESM) and lower-
secondary schools teachers data (RMSEA (90%CI) = .053 (.051–.056); CFI = .908; SRMR = .053; see Mod.5 in ESM)\(^2\). The CFI values still did not reach the suggested threshold of .95, but were closer to it than in the baseline model. The management factor correlated with the instructional factor at the level of .831 and .701, engagement with the instructional factor .904 and .833, and with the management factor .831 and .719 for primary and lower-secondary schools, respectively. In the model with a second-order factor the instructional, management and engagement factors loaded the second-order factor by .951, .874, .951, respectively for primary school teacher data and .902, .778, .924 for lower-secondary school teacher data. Factor loadings and their standard errors are presented in Table 3.

\(<\text{Table 3}>\)

**Measuring TSES invariance between primary and lower-secondary school teachers in Poland**

Only one analysis of TSES invariance was found in the literature. Klassen et al. (2009) tested the baseline three-factor oblique model of the TSES short form for a combined North American sample, an East Asian sample and a total of six samples included in their study. The TSES shows measurement invariance of form, factor loadings, factor variances and covariances across the three groups in the North American sample and across the two groups

\(^2\) The model with no error covariances allowed proved to have only slightly worse fit statistics: RMSEA (90%CI) = .050 (.047–.052); CFI = .902; SRMR = .045 (see Mod.6 in ESM) for primary school teacher data and RMSEA (90%CI) = .056 (.054–.058); CFI = .898; SRMR = .054 (see Mod.7 in ESM) for lower-secondary schools teacher data. Thus, allowing error covariances is not a threat to conclusions drawn here. For the model’s factor loadings and its robust standard errors please see Table 6 in ESM. The bi-factor model (Holzinger & Swineford, 1937) was also tested. Imposing a bi-factor structure on the data did not considerably improve fit either; RMSEA (90%CI) = .046 (.044–.049); CFI = .923; SRMR = .042 (see Mod.8 in ESM) for primary school teacher data and RMSEA (90%CI) = .050 (.047–.052); CFI = .925; SRMR = .046 (see Mod.9 in ESM) for lower-secondary school teacher data.
in the East Asian sample. Across all six groups, the model proved to be invariant in terms of factor structure and factor loadings. These results provide mixed support for model invariance regarding factor variances and covariances across all six settings studied. The authors did not, however, test model invariance across groups of elementary and middle versus secondary school teachers, which is covered in this article.

A dedicated analytical tool with which to assess measurement differences between groups is multi-group confirmatory factor analysis (MGCFA). All statistics in CFA based on $\chi^2$ (e.g. overall fit statistics, modification indices) are sensitive to sample size. Thus, although it is permissible to conduct MGCFA with significantly unequal group sizes, it is not recommended as it may lead to incorrect conclusions (Brown, 2006, p. 279). The ratio of the Polish joint sample to the U.S. sample is almost 18:1 (4465 to 255). This means that the Polish sample will contribute almost 18 times more to equal form $\chi^2$ than the U.S. sample. Moreover, and as already demonstrated, both samples are hardly balanced in terms of the distribution of teachers’ age, teaching experience, Grades taught, and subjects taught. The CFA model needed several adjustments to fit the Polish data well, and therefore the model structure no longer exactly matches the original one proposed by the scale originators. Taking these arguments into account, the invariance analysis across Polish and U.S. samples seems not worth the effort. However, it seems worth running MGCFA on the Polish data across primary and lower-secondary schools teachers. Especially since no such analyses have yet been published.

Several basic forms of factorial invariance have been listed in the literature (Meredith, 1993; Little & Slegers, 2005): (a) Configural factorial invariance ensures that in each group the same factorial structure is observed. (b) Metric (or pattern) invariance (also called weak factorial invariance) requires loadings to be equal across groups. If loadings are the same across groups then the scale is equally valid. (c) Strong factorial (or scalar) invariance
requires the loadings and indicators means (intercepts) to be equal across groups. (d) Strict factorial invariance requires all the above-mentioned conditions to be equal across groups, and in addition for the indicators’ residual variances to be equal across groups. If the indicators’ variances are equal, then the scale is equally reliable across groups. “These types of factorial invariance form a nested hierarchy primarily represented by increasing levels of cross-group equality constraints imposed on factor loading, item intercept, and residual variance parameters” (Gregorich, 2006, p. 3). Invariance was tested on the model with items 12 and 23 set to load also the instructional factor, and error covariances between 6 and 12, 3 and 23, 1 and 6 allowed. This model is the one advocated in this article as the one best fitting to the Polish data. The results are presented in Table 4.

<Table 4>

The results moderately support the measurement invariance of the tested model across the two groups. The p-values for the parent versus nested model differences tests based on $SB\chi^2$ (Satorra & Bentler, 2001) are below .01, suggesting that the parent model, i.e. the more complex one, with fewer constraints and more parameters to be estimated, fits data better. However, this is most likely due to the large sample size. On the other hand, RMSEA and SRMR suggest acceptable model fit in both groups, even with restrictions imposed on items’ loadings, intercepts and residual variances. Contrarily, CFI values for all models are below the assumed threshold for good fit. Also $\Delta$ CFI indicates that measurement invariance does not hold for Polish data. However, since the CFI compares the target and null model, the smaller loadings the less difference there is between a target and a null model. In such case CFI tends to be small, and seems not be that informative.
The model Mod.16 in fact does not test measurement invariance. Instead, the characteristics of the two groups are being tested; that is, are there any reasons to claim that neither the self-efficacy mean level nor its variability differ across primary and lower-secondary school teachers? Model Mod.16 resulted in a less satisfactory but still acceptable fit based on RMSEA values but not SRMR and CFI. In Mod.15 means were estimated in primary school teachers group for ISSE: -.158, CMSE: .140, SESE: .125 in standardized values in comparison to lower-secondary school teachers as reference group. Results of models Mod.15 and Mod.16 suggests small differences to both mean level and variability of sense of the self-efficacy between primary and lower-secondary school teachers.

Measurement invariance of the model with cross-loadings but no error covariances was also tested (see Table 7 in ESM). The pattern of results was similar.

**Discussion**

The results of the CFA on original in-service teacher data has been reported for the first time since the TSES was developed in 2001. The original scale form reflected in the data acceptable based on SRMR, however RMSEA and CFI were below the assumed thresholds of a good fit. It was hypothesized that CFI is not informative in the case of data from the U.S. and Poland studied in this article. Nevertheless, in several studies the TSES structure did replicate well, with CFI values over .95 (see Table 1).

The review of available CFA-based evidence from studies conducted during the 2001-2014 period generally support the three-factor structure of the teacher self-efficacy construct as well as TSES reliability. However, when administering the scale outside the U.S. several amendments were usually required. In not all studies summarized in Table 1 was the rationale for such amendments provided. Allowing for cross loadings and error covariances should never be driven purely by the desire to have the model fit the data better. Allowing for cross
loadings in the model examined on Polish data is justified, since two items in Polish translation overlap with the content of items representing the instructional factor. Allowing error covariances can be justified due to proximity of items in the questionnaire. With such justifications, the amendments implemented do not have a negative impact on the coherence of the self-efficacy theoretical construct. To support this claim models with cross loadings but without error covariances allowed were analyzed, and the bi-factor models. The results exhibit similar patterns to the models with error covariances allowed.

The evidence on the TSES factorial structure and reliability available in the literature is largely based on small convenience samples. In addition, most analyses were run on mixed-teacher samples, largely on pre-service teachers. This gap has been filled in by this article. It was concluded that on the basis of TSES analysis on large representative samples of Polish in-service primary and lower-secondary school teachers, TSES is a reliable measurement tool. Additionally, after some adjustments discussed above, the model fit to Polish data improved considerably.

Moderate support was found in the Polish data for the TSES measurement invariance of form, factor loadings, factor variances and covariances across primary and lower-secondary school teachers. This finding supports to some extent the claim that the TSES can be effectively used to measure self-efficacy among primary as well as lower-secondary school teachers. Measurement invariance between primary and lower-secondary school teachers has not been tested before this article.

The last important remark is that analysis of both the original U.S. data (used for TSES development back in 2001) and the Polish data reveals strong subscales correlations, which is a strong supporting evidence for the TSES as a measurement instrument for general self-efficacy beliefs assessment.
Acknowledgments

This article uses data from the studies co-financed from the European Social Fund under grants UDA.POKL.03.02.00-00-010/07-04 and UDA-POKL.03.02.00-00-001/13-00.

The author would like to thank Megan Tschannen-Moran for providing the original data set from the study on TSES development (Tschannen-Moran & Woolfolk Hoy, 2001).

Appendix

<Table 5>
References


planning: Dynamic testing, teaching’ experience and teachers’ sense of efficacy. Teaching
and Teacher Education, 28, 560–567.

Guilford Press.


Caprara, G. V., Barbaranelli, C., Borgogni, L., & Steca, P. (2003). Efficacy beliefs as
determinants of teachers’ job satisfaction. Journal of Educational Psychology, 95, 821–
832.

school staff’s and parents’ efficacy beliefs as determinants of attitudes toward school.

beliefs as determinants of job satisfaction and students’ academic achievement: A study at

Central Statistical Office (2012). Tabl. 3(238). In Education in 2011/12 School Year
(Statistical Information and Elaborations). Warsaw. Retrieved from Central Statistical
Office website:
http://stat.gov.pl/download/gfx/portalinformacyjny/pl/defaultaktualnosci/5488/1/10/1/calos
c_oiw_2011_2012.zip

Experimental Education, 60, 323–37.


Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square

with teacher engagement, job satisfaction, and emotional exhaustion. *Psychological
Reports, 114*, 68–77.

relationships between job satisfaction, sense of efficacy and teachers’ extra-role behavior.


Tobin, T. J., Muller, R. O., & Turner, L. M. (2006). Organizational learning and climate as

efficacy on teachers’ predictions of student success. *Teaching and Teacher Education, 21*,
299–314.


Effects on burnout, job satisfaction, illness, and quitting intentions. *Teaching and Teacher Education, 47*, 120–130.


Table 1. Summary of the CFA-based evidence of the TSES original factorial structure (fit statistics provided for the oblique confirmatory three-factor model, i.e. the baseline model)

<table>
<thead>
<tr>
<th>Study number</th>
<th>Article</th>
<th>n</th>
<th>Sampling type</th>
<th>Respondents type</th>
<th>Country of the scale administration</th>
<th>Scale form</th>
<th>ISSE</th>
<th>CMSE</th>
<th>SESE</th>
<th>α for unidimensional scale</th>
<th>χ²/df ratio</th>
<th>RMSEA</th>
<th>CFI</th>
<th>Subscale correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tschannen-Moran &amp; Woolfolk Hoy (2001)*</td>
<td>255</td>
<td>convenience</td>
<td>in-service preschool, elementary, middle/Jr. high, high school teachers</td>
<td>U.S.</td>
<td>long</td>
<td>.88</td>
<td>.87</td>
<td>.84</td>
<td>.93</td>
<td>2.2 (robust)</td>
<td>.069</td>
<td>.871</td>
<td>.60–.68</td>
</tr>
<tr>
<td>2</td>
<td>Çapa, Çakıroğlu, &amp; Sarıkaya (2005)</td>
<td>628</td>
<td>convenience</td>
<td>teacher students</td>
<td>Turkey</td>
<td>long</td>
<td>.86</td>
<td>.84</td>
<td>.82</td>
<td>.93</td>
<td>N/A</td>
<td>.065</td>
<td>.990</td>
<td>.74–.88</td>
</tr>
<tr>
<td>3</td>
<td>Gür, Çakıroğlu, &amp; Çapa Aydin (2012)</td>
<td>383</td>
<td>convenience</td>
<td>elementary school teachers</td>
<td>Turkey</td>
<td>long</td>
<td>.91</td>
<td>.90</td>
<td>.87</td>
<td>N/A</td>
<td>N/A</td>
<td>.080</td>
<td>.980</td>
<td>.87–.90</td>
</tr>
<tr>
<td>4</td>
<td>Mohamadi &amp; Asadzadeh</td>
<td>284</td>
<td>convenience</td>
<td>in-service high school teachers</td>
<td>Iran</td>
<td>short</td>
<td>.89</td>
<td>.91</td>
<td>.95</td>
<td>.84</td>
<td>1.76</td>
<td>.056</td>
<td>.980</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Study Reference</td>
<td>Sample Size</td>
<td>Population/Setting</td>
<td>Country</td>
<td>Short/Long</td>
<td>Effect Size</td>
<td>N/A</td>
<td>95% CI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-----------------</td>
<td>-------------</td>
<td>-------------------</td>
<td>---------</td>
<td>------------</td>
<td>-------------</td>
<td>-----</td>
<td>--------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Heneman, Kimball, &amp; Milanowski (2006)</td>
<td>1,075</td>
<td>elementary, middle and high school teachers (The Washoe County School District)</td>
<td>U.S.</td>
<td>short</td>
<td>.75–.90</td>
<td>N/A</td>
<td>6.62 .070 .960 .44–.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Wolters &amp; Daugherty (2007)</td>
<td>512</td>
<td>convenience through 12th-grade teachers</td>
<td>U.S.</td>
<td>long</td>
<td>N/A</td>
<td>N/A</td>
<td>5.62 .100 .860 N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Duffin, French, &amp; Patric (2012)</td>
<td>272</td>
<td>convenience undergraduate pre-service teachers</td>
<td>U.S.</td>
<td>long</td>
<td>.91 .91 .89 .96</td>
<td>4.11 .060 (SRMR) .960 .76–.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Bosma, Hessels, &amp; Resing (2012)</td>
<td>180</td>
<td>random at school level only elementary school teachers</td>
<td>Netherlands</td>
<td>long</td>
<td>.91 .90 .87 N/A</td>
<td>2.21 .070 .820 N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Klassen et al. (2009)</td>
<td>210</td>
<td>elementary &amp; middle school teachers</td>
<td>Canada</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>1.77 .061 .966</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>153</td>
<td>elementary &amp; middle school teachers</td>
<td>Cyprus</td>
<td>short</td>
<td>N/A</td>
<td>N/A</td>
<td>3.28 .105 .929</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>137</td>
<td>elementary &amp; middle school teachers</td>
<td>Korea</td>
<td>short</td>
<td>N/A</td>
<td>N/A</td>
<td>3.71 .134 .876</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>137</td>
<td></td>
<td>U.S.</td>
<td></td>
<td></td>
<td></td>
<td>1.38 .053 .974</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>255</td>
<td>Secondary school teachers</td>
<td>Canada</td>
<td>2.12</td>
<td>.066</td>
<td>.946</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-----</td>
<td>-----</td>
<td>---------------------------</td>
<td>--------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>247</td>
<td></td>
<td></td>
<td>Singapore</td>
<td>2.38</td>
<td>.075</td>
<td>.967</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Notes. N/A = not available; *Based on analysis conducted by the author of this paper on the original data, used for the scale development, provided by Megan Tschannen-Moran.*
Table 2. Factor loadings with robust (to non-normality) standard errors (in brackets) of the oblique confirmatory three-factor model ran on the original Tschannen-Moran and Woolfolk Hoy (2001) in-service 255 teachers data

<table>
<thead>
<tr>
<th>Item</th>
<th>Instruction</th>
<th>Item</th>
<th>Management</th>
<th>Item</th>
<th>Engagement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.699 (.049)</td>
<td>9</td>
<td>.820 (.031)</td>
<td>17</td>
<td>.771 (.036)</td>
</tr>
<tr>
<td>2</td>
<td>.739 (.039)</td>
<td>10</td>
<td>.692 (.044)</td>
<td>18</td>
<td>.698 (.041)</td>
</tr>
<tr>
<td>3</td>
<td>.689 (.049)</td>
<td>11</td>
<td>.784 (.040)</td>
<td>19</td>
<td>.684 (.044)</td>
</tr>
<tr>
<td>4</td>
<td>.721 (.041)</td>
<td>12</td>
<td>.748 (.036)</td>
<td>20</td>
<td>.611 (.049)</td>
</tr>
<tr>
<td>5</td>
<td>.647 (.048)</td>
<td>13</td>
<td>.771 (.036)</td>
<td>21</td>
<td>.700 (.037)</td>
</tr>
<tr>
<td>6</td>
<td>.658 (.045)</td>
<td>14</td>
<td>.736 (.041)</td>
<td>22</td>
<td>.625 (.043)</td>
</tr>
<tr>
<td>7</td>
<td>.630 (.046)</td>
<td>15</td>
<td>.490 (.064)</td>
<td>23</td>
<td>.537 (.053)</td>
</tr>
<tr>
<td>8</td>
<td>.599 (.052)</td>
<td>16</td>
<td>.501 (.058)</td>
<td>24</td>
<td>.540 (.062)</td>
</tr>
</tbody>
</table>

*Note.* Since results of confirmatory model are presented, blank cells indicate that loadings was forced to equal zero.
Table 3. Factor loadings with robust (to non-normality and teachers clustering within schools) standard errors (in brackets) of the oblique confirmatory three-factor model with error covariances between items 6 and 12, 3 and 23, 1 and 6 allowed, ran on Polish data

<table>
<thead>
<tr>
<th>Factor/item</th>
<th>Primary school teachers (2,083)</th>
<th>Lower-secondary school teachers (2,382)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instruction:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.767 (.017)</td>
<td>.745 (.014)</td>
</tr>
<tr>
<td>2</td>
<td>.753 (.020)</td>
<td>.735 (.015)</td>
</tr>
<tr>
<td>3</td>
<td>.727 (.017)</td>
<td>.695 (.019)</td>
</tr>
<tr>
<td>4</td>
<td>.789 (.019)</td>
<td>.776 (.014)</td>
</tr>
<tr>
<td>5</td>
<td>.637 (.028)</td>
<td>.611 (.018)</td>
</tr>
<tr>
<td>6</td>
<td>.748 (.017)</td>
<td>.706 (.017)</td>
</tr>
<tr>
<td>7</td>
<td>.696 (.025)</td>
<td>.667 (.017)</td>
</tr>
<tr>
<td>8</td>
<td>.721 (.026)</td>
<td>.712 (.014)</td>
</tr>
<tr>
<td>Management:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>.722 (.024)</td>
<td>.772 (.012)</td>
</tr>
<tr>
<td>10</td>
<td>.773 (.023)</td>
<td>.793 (.014)</td>
</tr>
<tr>
<td>11</td>
<td>.796 (.016)</td>
<td>.846 (.011)</td>
</tr>
<tr>
<td>12</td>
<td>.740 (.045)</td>
<td>.51 (.044)</td>
</tr>
<tr>
<td>13</td>
<td>.797 (.016)</td>
<td>.836 (.011)</td>
</tr>
<tr>
<td>14</td>
<td>.762 (.019)</td>
<td>.782 (.014)</td>
</tr>
<tr>
<td>15</td>
<td>.689 (.027)</td>
<td>.583 (.018)</td>
</tr>
<tr>
<td>16</td>
<td>.777 (.022)</td>
<td>.766 (.014)</td>
</tr>
<tr>
<td>Engagement:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>.715 (.024)</td>
<td>.686 (.020)</td>
</tr>
<tr>
<td>19</td>
<td>.749 (.023)</td>
<td>.661 (.020)</td>
</tr>
<tr>
<td>17</td>
<td>.694 (.020)</td>
<td>.617 (.024)</td>
</tr>
<tr>
<td>20</td>
<td>.746 (.019)</td>
<td>.685 (.018)</td>
</tr>
<tr>
<td>21</td>
<td>.711 (.021)</td>
<td>.651 (.020)</td>
</tr>
<tr>
<td>22</td>
<td>.640 (.025)</td>
<td>.539 (.027)</td>
</tr>
<tr>
<td>23</td>
<td>.588 (.084)</td>
<td>.143 (.086)</td>
</tr>
<tr>
<td>24</td>
<td>.642 (.026)</td>
<td>.558 (.026)</td>
</tr>
</tbody>
</table>
Note. Since results of confirmatory model are presented, blank cells indicate that loadings was forced to equal zero.
Table 4. Results of measurement invariance analysis of the oblique confirmatory three-factor model with cross-loadings and error covariances between items 6 and 12, 3 and 23, 1 and 6 allowed, ran on Polish data

<table>
<thead>
<tr>
<th>Invariance type</th>
<th>SBχ²</th>
<th>df</th>
<th>Scaling correction factor (c)</th>
<th>RMSEA (90% CI)</th>
<th>SRMR</th>
<th>CFI</th>
<th>SBχ²/df</th>
<th>Δ RMSEA</th>
<th>Δ CFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mod.10: Configural</td>
<td>3019.846</td>
<td>488</td>
<td>2.1168</td>
<td>0.048 (.047–.050)</td>
<td>0.048</td>
<td>0.916</td>
<td>6.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mod.11: Metric</td>
<td>3160.697</td>
<td>511</td>
<td>2.1063</td>
<td>0.048 (.047–.050)</td>
<td>0.060</td>
<td>0.912</td>
<td>6.19</td>
<td>0.000</td>
<td>-0.04*</td>
</tr>
<tr>
<td>Mod.12: Scalar</td>
<td>3624.548</td>
<td>522</td>
<td>2.0960</td>
<td>0.051 (.049–.053)</td>
<td>0.068</td>
<td>0.898</td>
<td>6.94</td>
<td>0.003</td>
<td>-0.014*</td>
</tr>
<tr>
<td>Mod.13: Strict</td>
<td>3716.497</td>
<td>556</td>
<td>2.1130</td>
<td>0.050 (.049–.052)</td>
<td>0.080</td>
<td>0.896</td>
<td>6.68</td>
<td>-0.001</td>
<td>-0.002</td>
</tr>
<tr>
<td>Mod.14</td>
<td>3749.811</td>
<td>559</td>
<td>2.1146</td>
<td>0.051 (.049–.052)</td>
<td>0.121</td>
<td>0.895</td>
<td>6.71</td>
<td>0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td>Mod.15</td>
<td>3799.774</td>
<td>562</td>
<td>2.1171</td>
<td>0.051 (.049–.052)</td>
<td>0.121</td>
<td>0.893</td>
<td>6.76</td>
<td>0.000</td>
<td>-0.002</td>
</tr>
<tr>
<td>Mod.16</td>
<td>3968.672</td>
<td>569</td>
<td>2.1215</td>
<td>0.052 (.050–.053)</td>
<td>0.153</td>
<td>0.888</td>
<td>6.97</td>
<td>0.001</td>
<td>-0.005*</td>
</tr>
</tbody>
</table>

Note. See detailed results in ESM. Mod.14) Strict with equality between factor covariances. Mod.15) Strict with equality between factor covariances and equal error covariances. Mod.16) Strict with equality between factor covariances, equal error covariances, and equal factor means and variances (models in all aspects exactly the same across groups). * Crossed threshold of Δ RMSEA > .007 or Δ CFI < -0.02, suggesting that hypothesis about measurement invariance should be rejected.
### Table 5. English TSES version and its Polish adaptation

**Instruction:**

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th>Polish</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>To what extent can you... use a variety of assessment strategies?</td>
<td>W jakim stopniu potrafisz... stosować zróżnicowane metody oceniania uczniów?</td>
</tr>
<tr>
<td>2</td>
<td>(...) provide an alternative explanation or example when students are confused?</td>
<td>(...) podać alternatywne wyjaśnienie lub przykład, kiedy uczniowie czegoś nie rozumieją?</td>
</tr>
<tr>
<td>3</td>
<td>(...) craft good questions for your students?</td>
<td>(...) ułożyć pytania inspirujące Twoich uczniów?</td>
</tr>
<tr>
<td>4</td>
<td>How well can you... implement alternative strategies in your classroom?</td>
<td>(...) stosować zróżnicowane metody nauczania?</td>
</tr>
<tr>
<td>5</td>
<td>(...) respond to difficult questions from your students?</td>
<td>(...) radzić sobie z trudnymi pytaniami uczniów?</td>
</tr>
<tr>
<td>6</td>
<td>How much can you do to adjust your lessons to the proper level for individual students?</td>
<td>(...) dostosować nauczanie do indywidualnego poziomu uczniów?</td>
</tr>
<tr>
<td>7</td>
<td>To what extent can you gauge student comprehension of what you have taught?</td>
<td>(...) ocenić na ile dobrze uczniowie rozumieją to, czego ich uczysz?</td>
</tr>
<tr>
<td>8</td>
<td>How well can you provide appropriate challenges for very capable students?</td>
<td>(...) dostarczyć odpowiednich wyzwań bardzo uzdolnionym uczniom?</td>
</tr>
</tbody>
</table>

**Management:**

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th>Polish</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>How much can you do to... control disruptive behavior in the classroom?</td>
<td>(...) zapanować przed zachowaniami, które przeszkadzają w prowadzeniu lekcji?</td>
</tr>
<tr>
<td>10</td>
<td>(...) get children to follow classroom rules?</td>
<td>(...) skłonić dzieci, by przestrzegały reguł obowiązujących w klasie?</td>
</tr>
<tr>
<td>11</td>
<td>(...) calm a student who is disruptive or noisy?</td>
<td>(...) uspokoić ucznia, który przeszkadza lub jest hałaśliwy?</td>
</tr>
<tr>
<td>12</td>
<td>How well can you... establish a classroom management system with each</td>
<td>(...) dostosować sposoby prowadzenia lekcji do różnych grup uczniów?</td>
</tr>
<tr>
<td>Item</td>
<td>Question</td>
<td>(Note)</td>
</tr>
<tr>
<td>------</td>
<td>----------</td>
<td>--------</td>
</tr>
<tr>
<td>13</td>
<td>(...) keep a few problem students from ruining an entire lesson?</td>
<td>(...) poradzić sobie z kilkoma trudnymi uczniami, żeby nie popsuli mi całej lekcji?</td>
</tr>
<tr>
<td>14</td>
<td>(...) respond to defiant students?</td>
<td>(...) poradzić sobie z buntującymi się uczniami?</td>
</tr>
<tr>
<td>15</td>
<td>To what extent can you make your expectation clear about student behavior?</td>
<td>(...) przekazać uczniom swoje oczekiwania wobec ich zachowania?</td>
</tr>
<tr>
<td>16</td>
<td>How well can you establish routines to keep activities running smoothly?</td>
<td>(...) wypracować w klasie zasady, które umożliwiają sprawne prowadzenie lekcji?</td>
</tr>
<tr>
<td>17</td>
<td>How much can you... do to get students to believe they can do well in schoolwork?</td>
<td>(...) sprawić, aby uczniowie uwierzyli, że mogą dobrze radzić sobie z nauką?</td>
</tr>
<tr>
<td>18</td>
<td>(...) do to help your students value learning?</td>
<td>(...) pokazać uczniom, że warto się uczyć?</td>
</tr>
<tr>
<td>19</td>
<td>(...) do to motivate students who show low interest in schoolwork?</td>
<td>(...) zmotywować uczniów, którzy nie są zainteresowani nauką?</td>
</tr>
<tr>
<td>20</td>
<td>(...) assist families in helping their children do well in school?</td>
<td>(...) pomóc rodzicom, aby ich dzieci dobrze radziły sobie w szkole?</td>
</tr>
<tr>
<td>21</td>
<td>(...) do to improve the understanding of a student who is failing?</td>
<td>(...) pomóc słabym uczniom zrozumieć materiał, z którym mają problemy?</td>
</tr>
<tr>
<td>22</td>
<td>(...) do to help your students think critically?</td>
<td>(...) pomóc swoim uczniom w krytycznym myśleniu?</td>
</tr>
<tr>
<td>23</td>
<td>(...) do to foster student creativity?</td>
<td>(...) rozwijać kreatywność uczniów?</td>
</tr>
<tr>
<td>24</td>
<td>(...) do to get through to the most difficult students?</td>
<td>(...) dotrzeć do bardzo trudnych uczniów?</td>
</tr>
</tbody>
</table>

**Note.** First four items in each subscale are used in the short form of the scale.