




European Journal of Educational Research


Volume 11, Issue 4, 2243 - 2263.

ISSN: 2165-8714

<http://www.eu-jer.com/>

Teaching and Student Evaluation Tasks: Cross-Cultural Adaptation, Psychometric Properties and Measurement Invariance of Work Tasks Motivation Scale for Teachers

Girum Tareke Zewude* 
University of Szeged,
HUNGARY/ Wollo University,
ETHIOPIA

Maria Hercz 
Eötvös Loránd University,
HUNGARY

Ngan Thi Ngoc Duong 
University of Szeged,
HUNGARY/ Kien Giang
University,
VIETNAM

Ferenc Pozsonyi 
Eötvös Loránd University,
HUNGARY

Received: April 11, 2022 ▪ Revised: July 30, 2022 ▪ Accepted: September 9, 2022

Abstract: The present research aimed to test an Amharic version of the multi-dimensional Work Task Motivation Scale for Teachers (WTMST), which measures the five pillars of university instructors' motivation toward teaching and student evaluation tasks based on self-determination theory (SDT). Therefore, the WTMST offers the first instrument to measure all five motivational elements, and today it is one of the most applicable instruments to assess teachers' motivation. An Amharic version of the WTMST for teaching and student evaluation tasks was adopted and assessed in large-scale data (N=1,117). Our findings demonstrate excellent reliability and construct validity (convergent, discriminant, divergent and factorial). Besides, the results of the model comparisons provided that out of the four theoretically competing models (single-order factor, correlated factor, higher-order factor and bi-factor models), the bi-factor model was the most-fitted one used for measurement invariance across various groups. Results also suggest that the factor structure of the WTMST for both teaching and student evaluation tasks demonstrate consistency across gender (men, women), university types (research, applied, and general university), age and experience in teaching. Therefore, the WTMST for teaching and student evaluation tasks may be valid in Ethiopian higher education settings.

Keywords: Student and teacher evaluation, work task motivation scale, wellbeing in higher education, cross-cultural adaptation.

To cite this article: Zewude, G. T., Hercz, M., Duong, T. N. N., & Pozsonyi, F. (2022). Teaching and student evaluation tasks: Cross-cultural adaptation, psychometric properties and measurement invariance of work tasks motivation scale for teachers. *European Journal of Educational Research*, 11(4), 2243-2263. <https://doi.org/10.12973/eu-jer.11.4.2243>

Introduction

The concept of teachers' motivation carries varying meanings and dimensions. For instance, Ryan and Deci (2000) defined motivation as energy, direction, and persistence, all the aspects of activation and intention. They further stated that motivation is essential in the real world because of its consequences. One of the most comprehensive empirically grounded theories of human motivation is the self-determination theory (SDT). It suggests that the scope of motivation is rather broad, uses concepts relevant to people, comprises a wide area of phenomena, is empirical-based, and has applicable values for human beings (Deci & Ryan, 2008a).

Nevertheless, SDT is a new approach for studying individuals' motivation and well-being. It has innumerable benefits in reaching fruitful results and could be exploited by parents, health care providers, religious leaders, managers, coaches, and teachers since it incorporates applicable theories and practices (Ryan & Deci, 2017). However, the predominance of the SDT is derived from the lack of any comprehensive theoretical motivational models that use different assessment tools in various contexts (Fernet et al., 2008; Gagné & Deci, 2005; Gagné et al., 2010).

The SDT distinguishes different types of motivation based on various goals or reasons that actuate rather than the amount of motivation (Deci & Ryan, 2008b; Fernet et al., 2008; Ryan & Deci, 2000). According to Fernet et al. (2008), the SDT has three broadly-known kinds of motivation from low to high: amotivation, extrinsic motivation, and intrinsic motivation. Gagné and Deci (2005) and Fernet et al. (2008) evidenced self-determined types of motivation is the most applicable theoretical model to assess teachers' work task motivation. *Amotivation* is the lack of motivation, and individuals overlook the consequences between their actions and their consequences; *external regulation* denotes

* Corresponding author:

Girum Tareke Zewude, Department of Education, Faculty of Humanities and Social Sciences, University of Szeged, Hungary. / Department of Psychology, Wollo University, Ethiopia. ✉ girumtareke27@gmail.com

controlled behaviour to avoid punishment or to get a reward; *introjected regulation* is an external demand derived from an internal representation, e.g. anxiety, shame, or guilt; in the cause of *identified regulation*, the action is owned or accepted as personally valuable because it matches with the subject's aims and values; and *intrinsic motivation* is the performance of an activity for the sake of innate gratification or satisfaction from the task itself.

Identified regulation and intrinsic motivation typically lead to positive consequences, whereas introjected regulation, external regulation, and amotivation lead to negative results (Gagne & Deci, 2005; Fernet et al., 2008). Also, self-determined types of motivation in work are associated with higher job satisfaction and better employees' well-being (Fernet et al., 2008). On the other hand, teachers' work task motivation is highly associated with burnout (Fernet et al., 2008) and teacher well-being (Collie et al., 2015). It is necessary to add that identified regulation and intrinsic motivation had a positive effect on the personal accomplishment of burnout inventory, whereas they are beneficial against emotional exhaustion and depersonalization (Fernet et al., 2008). In contrast, external regulation, introjected regulation and amotivation are positively associated with emotional exhaustion and depersonalization and negatively with personal accomplishment (Fernet et al., 2008). Consequently, it is more beneficial to assess the convergent (positive factors) and divergent (negative variables) validities of work task motivation with teacher well-being constructs (workload well-being, organizational and student interaction well-being; Collie et al., 2015), burnout (emotional exhaustion, depersonalization) and low personal accomplishment (Maslach et al., 1996).

Moreover, teachers on duty perform their tasks in the frame of three inseparable well-being factors. These are work-related well-being (e.g., students' evaluation, marking assignments and tests, working after hours), organizational well-being (e.g., support and recognition offered by administrators, relationship and communication with administrators, participation in administrative tasks, and established school rules and procedures), and student-interaction well-being (e.g., perceptions of student behaviour and motivation) (Collie et al., 2015).

In the Canadian cultural context, Fernet et al. (2008) examined elementary and high school teachers' psychometric properties of work task-based motivation based on the SDT. Also, the authors tested the construct validity of the work tasks motivation scale for teachers by incorporating six teachers' tasks (i.e., teaching, administrative, student evaluation, classroom preparation, classroom management and complementary tasks) using a correlated factor model (Fernet et al., 2008). Similarly, Fernet (2011) also ran a correlated factor model on the Work Role Motivation Scale for School Principals (WRMSSP) in the same cultural context. These two studies on work task motivation established and provided crucial psychometric information for current research. The authors also highlighted that future research should investigate multiple dimensions of teachers' motivation, examine construct validity, and employ multiple methods to minimize measurement biases (Fernet et al., 2008; Fernet, 2011).

In general, it could be stated that the study by Fernet et al. (2008) addresses WTMST based on the six teachers' main tasks (i.e., teaching, administrative, student evaluation, classroom preparation, classroom management and complementary tasks) using a correlated factor model. On the other hand, we found no research that compares the four computing models that are the single order factor, correlated factor, higher-order factor and bi-factor models, thus, we assume this research is able to provide a new comparative model and measurement invariance in the Ethiopian educational context and focus on two big teachers' tasks (i.e. teaching and student evaluation), giving new insight to university educators and other stakeholders working in the educational settings to improve teachers motivation.

The Objectives of the Present Study

The overall intention of our research is to contribute to the scientific usage of WTMST with a translation, adaption, psychometrical validation and equivalence measurement in Ethiopian education settings.

This investigation is justifiable for several reasons. First, a cross-cultural validation and adaptation of an instrument across groups could be proceeded after confirmed measurement equivalence or invariance (Davidov et al., 2014). Measuring variance or equivalence is essential for two main reasons: (a) the samples of the studies have different cultural backgrounds applying the same instrument, and (b) the data collected from different nations using different languages of the same instrument (Eremenco et al., 2005). Failure to establish measurement invariances hinders the sound interpretation of the data and the ability to demonstrate reliability and validity (Byrne & van de Vijver, 2010; Vandenberg & Lance, 2000). Therefore, some gaps are found in the literature to be investigated in other cultural contexts. Second, there have not been any validated studies on university teachers using model comparisons and measurement invariance concerning gender, age, university type, and experience in teaching in Ethiopia. Third, there has not been any research on the psychometric properties of the WTMST in the Amharic language or the African cultural context. Thus, we test Fernet et al.'s (2008) teaching and student evaluation tasks Ethiopian Amharic version translated for this study. Fourth, to test the construct validity of the WTMST, we run single and multi-confirmatory factor analysis using Ryan and Deci self-determination model assessed by Fernet et al. (2008). Fifth, we compare single, correlated, bi-factor, and higher-order factor models as those have been overlooked in earlier studies, and our examination is in accordance with recent methodological and analytical recommendations (Chen et al., 2006; Immekus & Imbrie, 2008; Liang & Luo, 2020; Stockdale et al., 2002; Wang et al., 2018). Sixth, as a final step, we select the best-

fitting model and perform further measurement invariance analysis across various groups to ensure the cross-cultural validation of the study (Chen et al., 2006; Eremenco et al., 2005).

In our questionnaire-based quantitative research, we formulate and test the following hypotheses:

Hypothesis 1: The five-factor structure of the WTMST is expected to be a reliable measurement in Ethiopian higher education settings.

Hypothesis 2: The five-factor structure of the WTMST (H1a-d; Fernet et al., 2008) show a good model fit using single-factor, correlated factor, bifactor, and higher-order confirmatory factor analysis (CFA) models with a group of university teachers.

Hypothesis 3: The WTMST show good convergent, discriminant and divergent validity.

Hypothesis 4: The WTMST measurement models show a good fit across gender, age, university type, and experience in teaching.

Hypothesis 5: The five WTMST of teaching tasks and student evaluation tasks are expected to be related positively.

Methodology

Research Design

The respondents of the survey completed the translated version of WTMST (Fernet et al., 2008) and its related measures, the Teacher Well-Being Scale (TWBS, Collie et al., 2015) and the Maslach Burnout Inventory (MBI, Maslach et al., 1996), in a cross-sectional survey from three Ethiopian public universities. Forty-seven questionnaires were excluded due to incompleteness before analysis; thus, the response rate was 95.3%. The completed data of the sample comprised (N=1,117) of which, n=835, 75% men, and n=282, 25% women) Ethiopian university teachers with a mean age of 31.1 years (SD=6.1 years).

Participants

Four hundred thirty-one (38.6%) participants were from the Gondar University (a research university), followed by 353 (31.6%) participants from the University of Woldia (a general university), and 333 (29.8%) participants from the university of Wollo (applied university); these institutions are set in the Amhara Regional State of Ethiopia. Table 1 illustrates the general socio-demographic information of all participants.

Table 1. Socio-Demographic Characteristics of the Participants (N=1,117)

Socio-Demographic Characteristics	Samples	
	N	Per cent
Gender		
Male	835	75
Female	282	25
Age		
25–35 years	695	62
36–45 years	365	33
46 years or above	57	5
University Type		
Research university	431	39
Applied university	333	30
General university	353	31
Experience in Teaching		
Below 5 years	411	37
6–10 years	293	26
11 years or above	413	37

Details of the Measurement Tools

The Work Task Motivation Scale for Teachers (WTMST). The WTMST by Fernet et al. (2008) used in this study seeks to answer the question, "Why are you teaching?" and "Why are you evaluating students?". The WTMST construct included 15 self-reported items on a 7-point scale, from 1 = does not correspond at all to 7= corresponds completely. The WTMST is based on SDT by Ryan and Deci (2000). It includes five sub-scales (Fernet et al., 2008), each with three items comprising the three types of motivation: intrinsic, extrinsic (identified, external, and introjected regulation), and amotivation (Fernet et al., 2008). It is a standardized, validated instrument for measuring teachers' motivation in the teaching context. As a result, intrinsic motivation, external regulation,

identified regulation, introjected regulation, and amotivation constructs are used in this study. In addition, the reliability of the five components of motivation was evaluated, and Cronbach's alpha values in the previous study ranged from 0.77 to 0.92 for all constructs. Finally, WTMST results provide excellent support for its psychometric properties (Fernet et al., 2008), and the scale was published by Fernet et al. (2008) in the *Journal of Career Assessment*. The construct reliability and validity information are presented in the result section. The adapted Amharic version of the WTMST measure was well established psychometric properties in the Ethiopian context (see the translation of the instrument in Appendix A).

Teacher Well-Being Scale (TWBS): The original version of the TWBS consisting of 16 items can be found in Collie et al. (2015), published in the *Journal of Psychoeducational Assessment*. Collie et al. (2015) showed that the teacher well-being construct has excellent internal and external validity and reliability (Collie, 2014; Collie et al., 2015). However, in this study, we used the Amharic version of TWBS (Zewude & Hercz, 2021, 2022). The authors performed various analyses to ensure the psychometric properties of the TWBS in the Ethiopian context. For further proofing of the construct validity, the TWBS construct validity was tested using the CFA model of the Ethiopian Amharic version, and the goodness of fit of the model shows: $\chi^2(101) = 455.63$, $p < 0.001$, TLI = 0.967, CFI = 0.972, RMSEA = 0.056, 95% CI [0.051, 0.061] in the Ethiopian Amharic language and culture. Therefore, the data had an acceptable fit following the recommended criteria: CFI, TLI ≥ 0.90 and RMSEA < 0.08 . This study found above 0.90 goodness of fit indices on TLI and CFI and RMSEA below 0.08; therefore, the finding of this study meets the acceptable recommended criteria of the CFA models. Besides the reliability of teacher well-being dimensions ranged from 0.80 to 0.2. The total scale was 0.92.

The Job Burnout: The Maslach Burnout Inventory (MBI) developed by Maslach et al. (1996) is the most widely used tool to assess job burnout. The MBI has 22 items listed in sub-dimensions *emotional exhaustion (EE)*, *personal accomplishment (PA)*, and *depersonalization (DP)*. Respondents rank items on a 7-point Likert scale, ranging from 0 = never to 6 = every day. In the current study, confirmatory factor analysis examined the model fit of the MBI scale with a robust maximum likelihood estimation method. Therefore, in this study the construct validity of the scale of confirmed the goodness of fit of the model: $\chi^2(206) = 1330.45$, $p < 0.001$, TLI = 0.931, CFI = 0.938, RMSEA = 0.070 [0.065, 0.074]. Moreover, Cronbach's alpha for the burnout subscales ranged from $\alpha = 0.885-0.942$, and the total scale reliability was $\alpha = 0.701$.

Data Analysis

The statistical analyses of this study were computed using IBM SPSS version 25.0 and Amos software version 26.0. Several multi-modal analyses were carried out to test this study's psychometric properties and measurement invariance of the Amharic version of WTMST, for example reliability (Cronbach's alpha and composite reliability), measurement invariance (configural, metric, scalar and residual), CFA (single and multi-group analysis), validity (divergent, convergent, discriminant), and structural equation modelling were employed (Davidov et al., 2018; Hair et al., 2019; Kline, 2016).

In addition, the absence of multi-collinearity was confirmed by examining the correlation matrices among the constructs, which should be less than 0.90, and by verifying the assumption of normality. Outliers of the constructs were also examined following the procedures of Hair et al. (2019), Kline (2016), Tabachnick and Fidell (2018). Values of $x \leq 2$ or $x \leq 4$ for skewness or kurtosis, respectively, indicate the normal distribution of data (Kim, 2013; Mishra et al., 2019). The skewness values are between -0.017 and -0.720 , and kurtosis scores range from -0.063 to -1.097 , which suggests all the constructs are relatively normally distributed (Table 2).

Table 2. Descriptive Statistics: Mean, Standard Deviation, Skewness and Kurtosis Scores

	Mean	Std. Deviation	Skewness	Kurtosis
Teacher Well-Being				
Workload Well-Being	27.151	5.322	-0.449	0.063
Organizational Well-Being	27.509	5.209	-0.509	0.069
Student Interaction Well-Being	17.046	4.408	-0.323	0.068
Teacher Well-Being	71.377	11.947	-0.044	-0.127
Teaching Task				
Intrinsic Motivation	12.05	2.349	-0.621	0.262
Identified Regulation	11.949	2.477	-0.772	0.886
External Regulation	11.604	2.764	-0.270	-0.011
Introjected Regulation	11.418	2.999	-0.372	-0.510
Amotivation	10.383	3.075	-0.245	-0.273
Teaching	57.405	8.439	-0.060	0.332
Student Evaluation Task				
Intrinsic Motivation	12.035	2.533	-0.528	0.678
Identified Regulation	12.138	3.135	0.042	1.049
External Regulation	8.203	3.468	0.094	-0.972
Introjected Regulation	8.087	3.730	0.316	-0.913
Amotivation	8.195	3.347	0.164	-0.655
Student Evaluation	54.965	8.006	-0.446	0.363
Burnout				
Depersonalization	16.760	5.801	-0.165	-0.911
Emotional Exhaustion	30.050	10.828	-0.720	-0.555
Personal accomplishment	18.846	10.250	0.255	-1.097
Burnout	65.657	10.673	-0.017	0.090

Reliability

The reliability scores of the current study were assessed by both Cronbach's alpha coefficient and composite reliability (CR). To test the accuracy of the reliability of the measurement of each construct following Cronbach's (1951) suggested guidelines are: the reliability value $\alpha \geq 0.9$ = excellent; the value range from α 0.89 to 0.80 = good; 0.79 to 0.70 = acceptable; 0.69 to 0.60 = questionable; α 0.59 to 0.50 = poor, and $0.50 > \alpha$ = unacceptable.

Convergent, Divergent, and Discriminant Validity

First, convergent and discriminant validity were assessed using the maximum shared variance (MSV) and the average variance extracted (AVE). The AVE values that exceed a threshold limit higher than 0.5 ($AVE > 0.05$) demonstrate good convergent validity. Moreover, factors whose MSV is lower than AVE are characterized by adequate discriminant validity (Hair et al., 2019). However, the researchers also expected the work task motivation to be positively associated with other similar positive variables (convergent validity) and negatively correlated with the opposite variables (divergent validity) (George & Mallery, 2020; Hair et al., 2019). As a result, Pearson's correlation coefficient was used to examine teacher well-being with the other relevant constructs (teacher well-being) to explore the convergent validity and teachers burnout to explore divergent validity. Schober et al. (2018) suggested the standard cut-points of correlational coefficients were: 0.90–1.00 = very strong, 0.70–0.89 = strong, 0.40–0.69 = moderate, 0.10–0.39 = weak, and 0.00–0.10 = negligible correlation.

Single and Multi-Modal Confirmatory Factorial Analysis (SMG-CFA)

The researchers used confirmatory factor analysis to test the construct validity of the WTMST (Fernet et al., 2008). The rationale used: firstly, the CFA test provides evidence of the validity of individual measures based on the overall fitness of model's and other evidence of construct validity (Hair et al., 2019). Secondly, the CFA analysis method was used because the selected variables previously hypothesized were theory-driven and empirically confirmed rather than derived from the data (Lei & Wu, 2007).

Model comparison can test further statistical analysis, address various assumptions, and reach a firm conclusion regarding the construct validity. Specifically, to select the best model, the following methods are recommended: first, by the method of information criteria, for example, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), second, by the best-fit measure using the maximum likelihood estimation (ML) such as Tucker-Lewis index (TLI), comparative fit index (CFI), the root mean square error of approximation (RMSEA) methods (Liang & Luo, 2020; Wang et al., 2018).

To achieve this, we have compared correlated factor models, the higher-order model, the single-factor model, and the bi-factor model of WTMST.

The details of the CFA models described in this study are as follows:

Model 1. First we examined the original inter-correlated five-factor WTMST of Fernet et al. (2008).

Model 2. The single-factor model included a single variable for each of the five core pillars of the work task motivation model to test a general factor against the dimensional factor structures (Immekus & Imbrie, 2008).

Model 3. Bi-factor model. Study the role of dimensions of work task motivation that are independent of the general (work task motivation) and to test both the general factor and dimensions of work task motivation as the primary concern of the issue (Chen et al., 2006, Wammerl et al., 2019). investigate model misfit due to fitting a unidimensional model to multi-dimensional and justify the creation of subscales (Immekus & Imbrie, 2008).

Model 4. Higher-order models were employed to examine the lower-order variables of WTMST are significantly associated with each other (Chen et al., 2006). In addition, higher-order factors also help reduce the number of path model relationships and provide means for handling collinearity among formative indicators (Johnson et al., 2011). In addition, bifactor and higher-order factor models are the two best alternative approaches for representing general constructs containing various highly correlated domains (Chen et al., 2006).

Regarding the fitness of indices in explaining our single and multi-modal CFA models, the researchers did not consider the chi-square test (χ^2) owing to the oversensitivity of sample size based on the suggestion of Barrett (2007) and Steiger (2007). Instead, we used the most globally reported goodness-of-fit indices: TLI, CFI, RMSEA and information criteria (AIC, BIC) are strongly recommended (Stockdale et al., 2002). The GFI is used to evaluate how justified the fitting model is, while TLI and CFI are for comparing the independent models to each model. Moreover, the TLI is free from the sample size, while the CFI estimates the sample model fit to the population (Stockdale et al., 2002). The following cut points are used for GFI, TLI and CFI, the recommended cut-points (cp) are the following: 1 = exact fit, 0.95 to 0.99 = close fit, 0.9 to 0.95 = acceptable fit, 0.85 to 0.9 = mediocre fit, and cp > 0.85 = poor fit (Hu & Bentler, 1999). Cut-points for RMSEA are as follows: 0 = exact fit, 0.01 to 0.049 = close fit, 0.05 to 0.079 = good fit, 0.08 to 0.09 = mediocre fit, and cp > 0.1 = poor fit (Hu & Bentler, 1999). For groups of 10 to 20, Hu and Bentler (1999) suggested RMSEA cut points of ≤ 0.08 for acceptable fit.

In addition to goodness-of-fit indices, the information criteria such as the AIC and BIC are the most appropriate for model comparison and valuable for selecting a good model (Hooper et al., 2008). For example, small values of AIC and BIC of the default model compared with saturated and independent models indicate good model fitting, and statistics also require a sample size of 200 and above to make their usage reliable (Hooper et al., 2008).

Measurement Invariance

Measurement invariance (MI) or equivalence refers to the unbiased measurement between two languages and cultural backgrounds using the same instrument (Eremenco et al., 2005), and it is needed to confirm comparative groups (culture, gender, age, education, etc.; Davidov et al., 2014). Such differences are detected by applying MI across various groups (Chen, 2007; Putnick & Bornstein, 2016). According to Byrne (2010), before the MI testing strategy begins, testing the CFA of each group separately for fit evidence is mandatory.

For the MI, we have tested the psychometric equivalence of the constructs across various groups using CFA (Putnick & Bornstein, 2016). The researchers tested the CFA models for the subgroups of gender, age, university type, and experience in teaching separately in the initial stage of this study. Good model fit is a prerequisite to testing MI (Byrne & van de Vijver, 2010). The researchers followed well-established scientific procedures using the four MI stages (Millsap, 2011; Putnick & Bornstein, 2016; Vandenberg & Lance, 2000). In stage 1, a configural invariance was conducted to establish a baseline model across groups without restriction, where the tested construct was the same across all groups (Vandenberg & Lance, 2000). In stage 2, the metric measurement invariance (MMI) was examined; the same constrained factorial loadings to the different groups responded similarly to indicators. In stage 3, scalar measurement invariance or strong invariance (SMI) was performed. In this test, the indicator intercepts and the factor loadings were constrained in the same way across groups. Finally, the residual measurement invariance or the strict invariance (RMI) was tested in the fourth stage. It refers to the similarity of item residuals of metric and scalar invariant items (Putnick & Bornstein, 2016). The MI of the present study four-sequential-staged analysis used single and multi-group CFA following Millsap (2011) and Putnick and Bornstein (2016) and arrived at the following recommendation criteria: Δ TLI, 0 = perfect and ≤ 0.01 = acceptable, Δ RMSEA, 0.015 for metric, scalar, and residual invariance (Chen, 2007; Putnick & Bornstein, 2016).

Findings

Reliability

The researchers evaluated the reliability of core dimensions of work task motivation for teaching and student evaluation tasks (amotivation, introjected regulation, external regulation, identified regulation, and intrinsic motivation) with the total sample ($N = 1,117$). As shown in Table 3, The reliability coefficients of the *WTMST of teaching tasks and student evaluation tasks* respectively were as follows (a) intrinsic motivation: $\alpha = 0.79, 0.82$; $CR = 0.80, 0.82$), (b) identified regulation $\alpha = 0.84, 0.91$; $CR = 0.84, 0.91$; (c) external regulation $\alpha = 0.83, 0.92$; $CR = 0.83, 0.92$; (d) introjected regulation $\alpha = 0.90, 0.94$; $CR = 0.90, 0.95$; and (e) amotivation $\alpha = 0.80, 0.89$; $CR = 0.81, 0.89$. Besides, the other related variables used for convergent (teacher well-being) and divergent validity (burnout) demonstrated acceptable reliability. This study showed that the five dimensions of teachers' motivation in both tasks were characterized by high reliability.

Convergent, Divergent and Discriminant Validity

In this study, we evaluated the validity of teachers' work task motivation in multi-dimensions for teaching and student evaluation tasks based on their respective scores in AVE and MSV (see Table 3). This study found that all five motivation constructs of teaching and student evaluation tasks have a good convergent validity ($AVE > 0.05$), implying that the corresponding items are composed of the core factors with acceptable correlation. The reason is that each item explains the latent constructs in each factor. Three methods were employed to test the discriminant validity of the work task motivation of teachers for both teaching and student evaluation tasks.

First, the constructs were tested since their AVE values were higher than their MSV (see Table 3). Hence, the result of this study indicated that the sub-constructs of the motivation scale (teaching tasks) AVE were greater than MSV and presented their consecutive result as follows (a) intrinsic motivation ($AVE = 0.57 > MSV = 0.08$), (b) identified regulation ($AVE = 0.64 > MSV = 0.23$), (c) external regulation ($AVE = 0.62 > MSV = 0.16$), (d) introjected regulation ($AVE = 0.74 > MSV = 0.16$), and e) amotivation ($AVE = 0.58 > MSV = 0.10$). Regarding the WTMST scale of student evaluation tasks are as follows: (a) for introjected regulation ($AVE = 0.85 > MSV = 0.25$), (b) external regulation ($AVE = 0.79 > MSV = 0.27$), (c) identified regulation ($AVE = 0.77 > MSV = 0.05$), (d) amotivation ($AVE = 0.72 > MSV = 0.27$), followed by intrinsic motivation ($AVE = 0.61 > MSV = 0.05$).

Second, we compared the AVE with squared inter-item correlations within the construct (see Table 3) to assess discriminant validity to see whether the AVE is higher than squared correlations (Hair et al., 2019). Thus, based on the suggested criteria, AVE is higher for all the constructs in this study than the squared correlation of each construct, which suggests that each factor's variance is better explained by the corresponding items that mainly load on each factor.

Third, to demonstrate the convergent validity of the work task motivation of teachers' construct, the researchers tested its association with other relevant positive psychology constructs (teacher well-being). The result revealed that the teacher motivation construct for both tasks (teaching and student evaluation tasks) have a statistically and positively significant relationship with teacher well-being ($r = 0.1328, p < 0.01$, and $r = 0.082, p < 0.01$, respectively). In addition, the researchers tested teachers' work task motivation divergent validity by measuring their relationship with negative psychological constructs (burnout). The result of divergent validity showed that teachers' motivation for both student evaluation tasks was negatively and significantly correlated to a negative variable of burnout ($r = -0.122, p < 0.05$). However, it was positively associated with WTMST of teaching task ($r = 0.106, p < 0.05$) (see Table 4).

Based on the results obtained, we can confidently conclude that the work task motivation of teachers in both teaching and student evaluation tasks in the Ethiopian higher educational settings meets the requirements of convergent, discriminant, and divergent validity. The results further suggest that the instruments are applicable.

Table 3. Reliability and Validity Indices of the Work Task Motivation Scale for Teachers (WTMST) for Teaching and Student Evaluation Tasks

Models	α (>.70*)	CR	AVE (>.50*)	MSV	Squared correlation					
					IM	IR	ER	INTR	AM	
IM	0.79	0.80	0.57	0.08	1					
IR	0.83	0.84	0.64	0.23	0.06	1				
ER	0.82	0.83	0.62	0.16	0.08	0.23	1			
ITR	0.90	0.90	0.74	0.16	0.08	0.08	0.16	1		
AM	0.80	0.81	0.58	0.10	0.01	0.05	0.10	0.02	1	
IM	0.82	0.83	0.61	0.05	1					
IR	0.91	0.91	0.77	0.05	0.23	1				
ER	0.92	0.92	0.79	0.27	-0.20	-0.23	1			
ITR	0.94	0.95	0.85	0.25	-0.12	-0.20	0.52	1		
AM	0.89	0.89	0.72	0.27	-0.13	-0.15	0.50	0.47	1	

Note: Asterisk indicates a global rule of thumb of an acceptable level of validity and reliability based on the recommendation of Hair et al. (2019) and Kline (2016).

Legend: α = Cronbach's alpha; AM=amotivation, AVE=average variance extracted, CR=composite reliability, ER=external regulation, IM=intrinsic motivation, ITR=introjected regulation=identified regulation, MSV = maximum shared variance

Table 4. Pearson-Correlations of WTMST Scores and Related Measures

Variables	Convergent Validity	
	Work Task Motivation Scale for Teachers	
	Teaching task	Student evaluation task
Teacher well-being	0.132**	0.082**
	Divergent validity	
Burnout	0.106*	-0.122*

Note: *p < .05; **p < .01. Table 4 indicates convergent and divergent validity indices for WTMST factors

Confirmatory Factor Analysis: Model Comparison and Evaluation

The researchers examined the four competing models (single-factor model, correlated factor model, bi-factor model, and higher-order-factor model) for teachers' motivation for teaching and student evaluation tasks. We have selected the most appropriate model fit for further test measurement invariance (equivalence). One of the best criteria for selecting the best model is the highest goodness of fit (GFI, TLI, CFI and RMSEA) and the lowest AIC and BIC. As a result, the current study showed the highest GFI, TLI, CFI, RMSEA and smaller values of AIC and BIC of the bi-factor model on teaching and student evaluation tasks of work task motivation (see Table 5) than other models. Figures 4 (a–d) illustrated the four competitive models of the WTMST for teaching and student evaluation tasks.

Table 5. Comparison of Fit Indices in the Four Competitive Models for WTMST for both Teaching and Student Evaluation Tasks

Models	χ^2 (df)	p-value	χ^2/df	GFI	TLI	CFI	RMSEA (95%) CI	AIC	BIC
Single-Order Factor	4541.85 (90)	0.001	50.46	0.612	0.306	0.405	0.211 [0.205, 0.216]	4601.85	4752.40
Correlated Factor	243.18 (80)	0.001	3.04	0.972	0.971	0.978	0.043 [0.037, 0.049]	323.18	523.92
Bi-Factor	189.74(75)	0.001	2.53	0.977	0.979	0.985	0.037 [0.031, 0.044]	279.74	505.56
Higher-Order Factor	289.17 (85)	0.001	3.40	0.965	0.966	0.973	0.046 [0.041, 0.052]	359.17	534.81
Single-Order Factor	7153.74 (90)	0.001	79.49	0.516	0.313	0.411	0.265 [0.260, 0.270]	7213.74	7364.30
Correlated Factor	286.59 (80)	0.001	3.58	0.967	0.977	0.983	0.048 [0.042, 0.054]	366.59	567.24
Bi-Factor	202.72 (75)	0.001	2.67	0.977	0.985	0.989	0.039 [0.032, 0.045]	290.72	511.53
Higher-Order Factor	320.04 (85)	0.001	3.77	0.948	0.976	0.980	0.050 [0.044, 0.056]	390.04	565.68

Note: $N=1,117$, $p < 0.001$. AIC = Akaike information criterion. BIC= Bayesian information criterion; CFA=confirmatory factor analysis; df = degree of freedom; CFI = comparative fit index, GFI=goodness of fit; RMSEA = root mean squared error of approximation; TKI = Tucker-Lewis's index; χ^2 = chi-square.

As Table 5 demonstrates, except the single-factor model, three of the competing models, namely the correlated factor, the bi-factor, and the higher-order factor models of WTMST for both teaching and student evaluation tasks showed an excellent model fit. Regarding the best model fit among the four competing models both teaching, and student evaluation tasks showed best in the bi-factor model used for further measurement invariance analysis. The bi-factor, correlated factor, and the higher order model results of WTMST in teaching tasks found that, $\chi^2 (75) = 189.74, p < 0.001, \chi^2/df = 3.04, GFI = 0.977, TLI = 0.979, CFI = 0.985, RMSEA = 0.037, 95\% CI [0.031, 0.044], AIC=279.74, BIC=505.56; \chi^2 (80) = 243.18, p < 0.001, \chi^2/df = 2.53, GFI = 0.972, TLI = 0.971, CFI = 0.978, RMSEA = 0.043, 95\% CI [0.037, 0.049], AIC=323.18, BIC=523.92; and \chi^2 (85) = 289.17, p < 0.001, \chi^2/df = 3.40, GFI = 0.965, TLI = 0.966, CFI = 0.973, RMSEA = 0.046, 95\% CI [0.041, 0.052], AIC=359.17, BIC=534.81, respectively.$

Besides, the WTMST of three best fitted competitive models of this study in student evaluation tasks found that, $\chi^2 (76) = 202.22, p < 0.0016, \chi^2/df = 2.67, GFI = 0.977, TLI = 0.985, CFI = 0.989, RMSEA = 0.039, 95\% CI [0.032, 0.045], AIC=290.72, BIC= 511.53$ for bi-factor model; $\chi^2 (80) = 286.59, p < 0.001, \chi^2/df = 3.58, GFI = 0.967, TLI = 0.977, CFI = 0.983, RMSEA = 0.048, 95\% CI [0.042, 0.054], AIC = 366.59, BIC=567.24$ for correlated factor model; and $\chi^2 (85) = 320.04, p < 0.001, \chi^2/df = 3.77, GFI = 0.948, TLI = 0.976, CFI = 0.980, RMSEA = 0.050, 95\% CI [0.044, 0.056], AIC = 390.04, BIC = 565.68$ for higher-order factor model. Figure 1 illustrates the four competing models of WTMST of teaching and student evaluation tasks presented below.

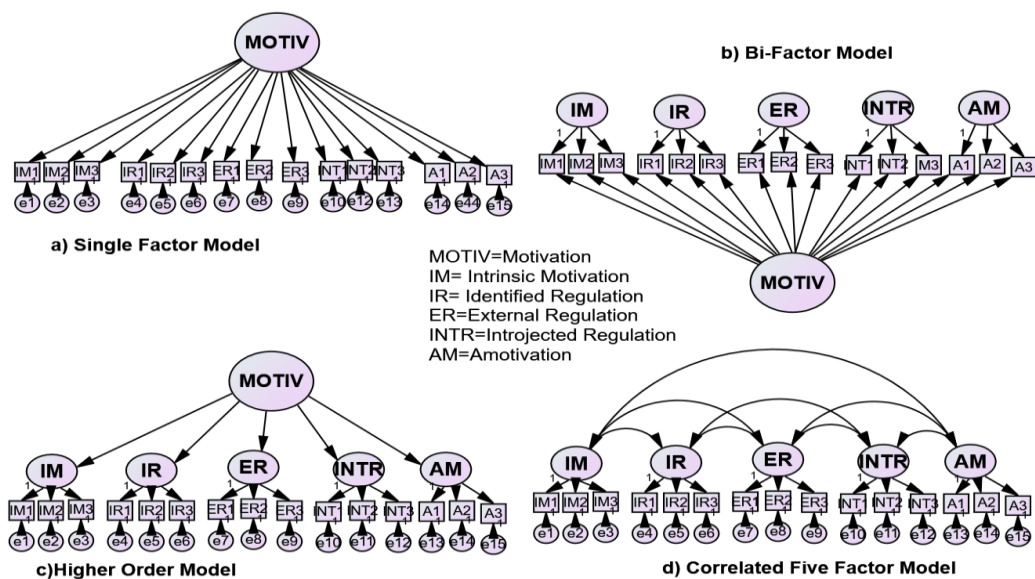


Figure 1. The Four Competing CFA Models of the Work Task Motivation Scales for Teachers for both Teaching and Student Evaluation Tasks

Note: (a) single-factor model, (b) bi-factor model, (c) Higher-order factor model, and (d) five-factor correlated model

The bi-factor model of the WTMST in teaching and student evaluation tasks showed the highest global fitness indices (GFI, TLI, CFI, and RMSEA) and the lowest in AIC and BIC compared with the other three comparison CFA models. On the other hand, higher-order and correlated factor models showed relatively congruent results, and the single-factor model showed the worst fitness of indices, and it was rejected due to its poor standard global cut-off points. The poor fitness of indices of the single-factor model implies that the WTMST for both teaching and student evaluation tasks are multi-dimensional constructs. Hence, the WTMST of teaching and student evaluation tasks show the best fit in this study in the three model comparisons: the bi-factor model, three-factor model, and higher-order factor model.

Additionally, the factor loadings of the three fitted models (bifactor, correlated and higher-order factor models) were high, and the factor loadings showed significance in all five dimensions ($p < 0.001$). All the 15 items of the WTMST in teaching and student evaluation tasks of the three competing models factor loadings higher than 0.40 and significant (see Table 6).

To sum up, the model comparison of this study valued these goodness-of-fit indices (GFI, TLI and $CFI > 0.90$, and $RMSEA < 0.08$), and AIC and BIC lowest score indicated that among the four competing models the bifactor model was the best-fitted used for further measurement invariance analysis to ensure and give a complete picture of cross-cultural validation.

Table 6 Standardized Factor Loadings for the Four Computing Models on the Work Task Motivation Scale for Teachers

WTMST for both teaching and student evaluation tasks, respectively

Sub Dimension	Items	Correlated Factor Model							
		Single actor Model	Bi- Factor Model	Higher-order Factor	IM	IR	ER	ITR	AM
Intrinsic motivation	IM1	0.33 (0.16)	0.82 (0.79)	0.84 (0.83)	0.84 (0.83)				
	IM2	0.43 (0.11)	0.53 (0.70)	0.66 (0.72)	0.66 (0.72)				
	IM3	0.29 (0.12)	0.71 (0.78)	0.74 (0.79)	0.75 (0.80)				
Identified regulation	IR1	0.58 (0.24)	0.55 (0.76)	0.76 (0.80)		0.76 (0.80)			
	IR2	0.56 (0.24)	0.71 (0.92)	0.85 (0.94)		0.85 (0.94)			
	IR3	0.51 (0.22)	0.66 (0.85)	0.78 (0.88)		0.78 (0.88)			
External regulation	ER1	0.60 (-0.60)	0.45 (-0.48)	0.72 (0.89)			0.72(0.89)		
	ER2	0.71 (-0.54)	0.47 (-0.59)	0.85 (0.87)			0.85(0.87)		
	ER3	0.67 (-0.56)	0.42 (-0.69)	0.79 (0.90)			0.79(0.90)		
Introjected regulation	ITR1	0.59 (-0.91)	0.69 (-0.58)	0.84 (0.90)				0.84(0.90)	
	ITR2	0.56 (-0.89)	0.80 (-0.71)	0.89 (0.93)				0.89(0.93)	
	ITR3	0.57 (-0.91)	0.74 (-0.74)	0.85 (0.95)				0.85(0.95)	
Amotivation	AM1	0.24 (-0.50)	0.63 (-0.56)	0.66 (0.82)					0.66 (0.82)
	AM2	0.31 (-0.47)	0.79 (-0.68)	0.85 (0.86)					0.84(0.86)
	AM3	0.26 (-0.49)	0.73 (-0.70)	0.77 (0.87)					0.77(0.88)

Note: $N=1,117$, $p < 0.001$. The factor loadings of all five dimensions of WTMST

Measurement Invariance (MI) WTMST for Both Teaching and Student Evaluation Tasks

The researchers considered the following cut-points as the most fitting. MI: $\Delta\text{CFI} = 0.02$ and $\Delta\text{RMSEA} = 0.03$ for the metric invariance test and $\Delta\text{CFI} = 0.01$ and $\Delta\text{RMSEA} \leq 0.015$ for the scalar and residual tests (Putnick & Bornstein, 2016). The teaching and student evaluation tasks of WTMST were constructed independently across groups, and the model shows a good fit. Single CFA was performed before testing MI (see Table 7). Figure 2 clearly illustrates the measurement invariance across groups of the teaching and student evaluation tasks of WTMST.

Measurement Invariance (MI) of WTMST for Teaching Tasks

Gender. Regarding to gender of university teachers with respect to student evaluation tasks of WTMST, men as compared women showed the best model fit to the data, ($\chi^2(75) = 161.86$, $p < 0.001$, $\chi^2/\text{df} = 2.13$, $\text{TLI} = 0.987$, $\text{CFI} = 0.990$, $\text{RMSEA} = 0.037$, $95\% \text{ CI} [0.029, 0.045]$, and ($\chi^2(75) = 109.04$, $p < 0.01$, $\chi^2/\text{df} = 1.85$, $\text{TLI} = 0.985$, $\text{CFI} = 0.989$, $\text{RMSEA} = 0.039$, $95\% \text{ CI} [0.021, 0.055]$) respectively. Following the single CFA analysis, the MI result is presented as follows. The CMI test of student evaluation tasks yielded sufficient fit to the data at the bi-factor model regarding to the gender, $\chi^2(152) = 270.99$, $p < 0.001$, $\chi^2/\text{df} = 1.78$, $\text{TLI} = 0.986$, $\text{CFI} = 0.990$, $\text{RMSEA} = 0.026 [0.021, 0.032]$ for WTMST.

As displayed in Table 7, for the metric measurement invariance result revealed: $\Delta\chi^2(10) = 10.82$, $\Delta\text{TLI} = -0.001$, $\Delta\text{CFI} = 0.000$, $\Delta\text{RMSEA} = 0.000$, $95\% \text{ CI} [0.000- 0.001]$. This result indicates that the metric invariance test scores across gender were invariant. The items used to estimate the factorial loadings were the same for the two groups of the student evaluation tasks. The SMI was performed based on comparing metric and scalar measurement invariance. After computing the MMI, the more restrictive model constrained the item intercepts. The SMI factor loadings were calculated, which showed an acceptable fit for the motivation model of the teaching tasks: $\Delta\chi^2(15) = 11.01$, $\Delta\text{TLI} = -0.002$, $\Delta\text{CFI} = 0.000$, $\Delta\text{RMSEA} = 0.002$, $95\% \text{ CI} [0.002, 0.002]$. It can be deduced that the item intercept and the factorial loadings are the same. The residual invariance (constrained item intercept, factor loadings and strict measurement invariance) fitness indices across gender found that, $\Delta\chi^2(15) = 42.67$, $\Delta\text{TLI} = 0.004$, $\Delta\text{CFI} = 0.003$, $\Delta\text{RMSEA} = 0.002$, $95\% \text{ CI} [0.002, -0.001]$.

Age. The age of university teachers (25 to 35 years, 36 to 45 years, 46 years and above), showed the best model fit to the data, ($\chi^2(75) = 179.28$, $p < .001$, $\chi^2/\text{df} = 2.36$, $\text{TLI} = 0.984$, $\text{CFI} = 0.988$, $\text{RMSEA} = 0.044 [0.036, 0.053]$, and ($\chi^2(75) = 152.54$, $p < 0.001$, $\chi^2/\text{df} = 2.01$, $\text{TLI} = 0.963$, $\text{CFI} = 0.973$, $\text{RMSEA} = 0.053 [0.040, 0.065]$, and ($\chi^2(75) = 89.79$, $p > 0.05$, $\chi^2/\text{df} = 1.20$, $\text{TLI} = 0.948$, $\text{CFI} = 0.963$, $\text{RMSEA} = 0.059 [0.000, 0.091]$. respectively. Still, the rest two categories showed an excellent model fit and the measurement invariance or equivalence was not affected (see Table 7).

We examined MI group differences using the bifactor model of the WTMST for student evaluation tasks across age categories. The CMI test yielded sufficient fit to the data at the bi-factor model, ($\chi^2(228) = 425.18$, $p < 0.001$, $\chi^2/\text{df} = 1.86$, $\text{TLI} = 0.977$, $\text{CFI} = 0.984$, $\text{RMSEA} = 0.028 [0.024, 0.032]$ for student evaluation task. The metric MI were $\Delta\chi^2(20) = 39.63$, $\Delta\text{TLI} = 0.000$, $\Delta\text{CFI} = 0.002$, $\Delta\text{RMSEA} = 0.000$, $95\% \text{ CI} [0.000, 0.000]$. However, the scalar and residual MI of the WTMST of student evaluation tasks across university groups was not fitted with global cut off points.

University Type. The Bi-factor motivation model of student evaluation tasks showed an excellent fit to the data regarding university type (research, applied, or general): $\chi^2(75) = 135.65$, $P < .001$, $\chi^2/\text{df} = 1.81$, $\text{TLI} = 0.983$, $\text{CFI} = 0.988$, $\text{RMSEA} = 0.043$, $95\% \text{ CI} [0.031, 0.055]$, ($\chi^2(75) = 140.69$, $P < .001$, $\chi^2/\text{df} = 1.88$, $\text{TLI} = 0.972$, $\text{CFI} = 0.980$, $\text{RMSEA} = 0.051$, $95\% \text{ CI} [0.038, 0.064]$, and ($\chi^2(75) = 138.98$, $P < .001$, $\chi^2/\text{df} = 1.83$, $\text{TLI} = 0.979$, $\text{CFI} = 0.985$, $\text{RMSEA} = 0.049$, $95\% \text{ CI} [0.036, 0.061]$, respectively. The WTMST of student evaluation tasks at the bi-factor model also showed an excellent fit to the data regarding university type at all four MI stages and presented the result as follows (see Table 7). Configural = $\chi^2(228) = 408.35$, $\chi^2/\text{df} = 1.79$, $\text{TLI} = 0.980$, $\text{CFI} = 0.985$, $\text{RMSEA} = 0.027$, $95\% \text{ CI} [0.022, 0.031]$, metric = $\Delta\chi^2(20) = 48.62$, $\Delta\text{TLI} = 0.002$, $\Delta\text{CFI} = 0.002$, $\Delta\text{RMSEA} = -0.001$, $95\% \text{ CI} [-0.002, -0.001]$, scalar = $\Delta\chi^2(89) = 153.92$, $\Delta\text{TLI} = 0.009$, $\Delta\text{CFI} = 0.010$, $\Delta\text{RMSEA} = -0.005$, $95\% \text{ CI} [-0.005, 0.004]$, and residual = $\Delta\chi^2(24) = 215.78$, $\Delta\text{TLI} = 0.014$, $\Delta\text{CFI} = 0.016$, $\Delta\text{RMSEA} = -0.007$, $95\% \text{ CI} [-0.007, -0.005]$.

Experience in teaching. University teachers' years of teaching experience (fewer than 5 years, 6 to 10 years, and 11 years or more) showed an excellent fit to the data (see Table 7), as did MI on the configural, metric, scalar, and residual tests. In addition, the strict model (residual) was achieved, and all item loadings, intercepts, and residual variances were equivalent or equal across the three levels of experience in teaching.

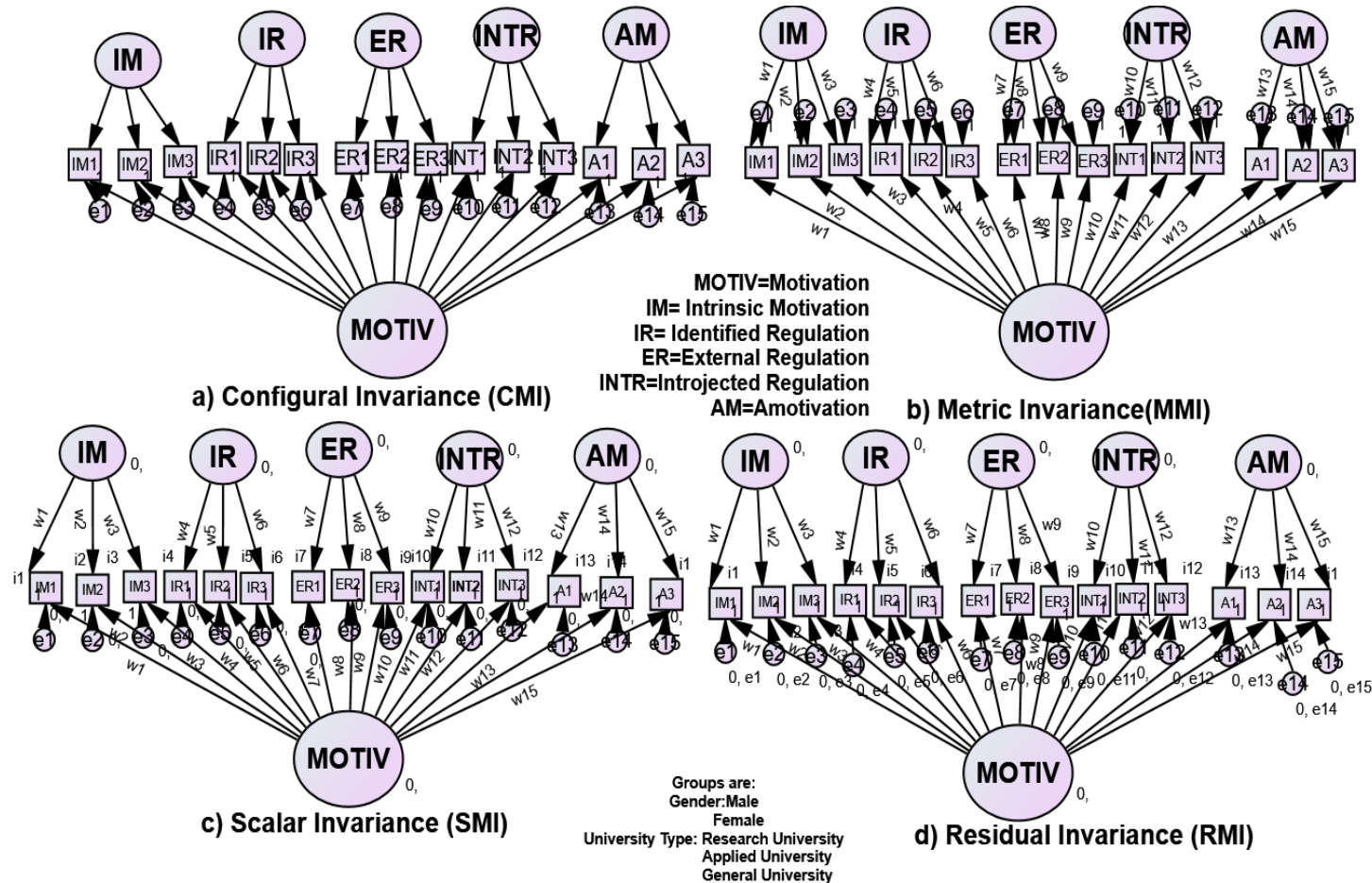


Figure 2. Measurement Invariance of The Bi-factor of Teacher's Motivation Model for both Teaching and Student Evaluation Tasks

Note (a) the configural MI (the same reference), (b) the metric MI (the same factor loadings), (c) the scalar MI: the equivalence of intercepts, and (d) the residual MI (equality or similarity of errors). Again, the bi-factor model was used for MI.

Table 7. Single-Factor Group and Measurement Invariance of The WTMST (Teaching and Student Evaluation Tasks) across Various Groups

Work Tasks Motivation Scale for Teachers (WTMST)- Teaching Tasks										
Models	Steps of MI	χ^2 (df)	χ^2/df	TLI	CFI	RMSEA 95 % CI	Comparison	Δ TLI	Δ CFI	Δ RMSEA (95 % CI)
Gender										
MI	Male (n=835)	170.55(75)*	2.27	0.976	0.983	0.039 [0.031, 0.047]				
	Female (n=282)	114.06(75)*	1.52	0.971	0.980	0.043 [0.026, 0.058]				
	M1: Configural	284.71(150)*	1.90	0.975	0.982	0.028 [0.023, 0.033]	—	—	—	—
	M2: Metric	292.97 (160)*	1.83	0.977	0.982	0.027 [0.022, 0.032]	M1	-0.002	0.000	0.001 [0.001, 0.001]
	M3: Scalar	319.26 (175)*	1.82	0.977	0.981	0.027 [0.022, 0.032]	M2	-0.000	0.001	0.000 [0.000, 0.000]
	M4: Residual	353.01(190)*	1.86	0.976	0.978	0.028 [0.023, 0.032]	M3	0.001	0.003	-0.001 [-0.000, 0.000]
University Type										
MI	Research University (n=431)	150.42(75)*	2.00	0.965	0.975	0.048 [0.037, 0.060]				
	Applied University (n=333)	96.03(75)*	1.28	0.987	0.991	0.029 [0.000, 0.045]				
	General University (n=353)	128.58(75)*	1.71	0.966	0.976	0.045 [0.031, 0.058]				
	M1: Configural	375.03 (225)*	1.67	0.972	0.980	0.024 [0.020, 0.029]	—	—	—	—
	M2: Metric	413.84(1.69)*	1.69	0.971	0.978	0.025 [0.021, 0.029]	M1	0.001	0.002	-0.001 [-0.001, 0.000]
	M3: Scalar	480.06 (275)*	1.74	0.969	0.973	0.026 [0.022, 0.030]	M2	0.002	0.005	-0.001 [-0.001, -0.001]
M4: Residual	533.51 (305)*	1.75	0.969	0.970	0.026 [0.022, 0.030]	M3	0.001	0.003	0.003 [0.000, -0.000]	
Age										
MI	25-35 years (n=695)	179.28(76)*	2.06	0.979	0.985	0.039 [0.030, 0.048]				
	36-45 years (n=365)	118.12(76)*	1.57	0.970	0.978	0.040 [0.025, 0.053]				
	46+ years(n=57)	114.30(75)*	1.52	0.891	0.922	0.097 [0.059, 0.131]				
	M1: Configural	390.00 (225)*	1.73	0.970	0.979	0.026 [0.021, 0.030]	—	—	—	—
	M2: Metric	416.23 (245)*	1.70	0.971	0.978	0.025 [0.021, 0.029]	M1	-0.001	0.001	0.001 [0.000, 0.001]
	M3: Scalar	607.91(248)*	2.45	0.936	0.949	0.036 [0.032, 0.040]	M2	0.035	0.029	-0.011 [-0.011, -0.011]
M4: Residual	869.65(305)*	2.85	0.918	0.921	0.041 [0.038, 0.044]	M3	0.018	0.028	-0.005 [-0.008, -0.004]	
Experience in Teaching										
MI	Below 5 years (n=411)	186.54 (75)*	2.49	0.948	0.963	0.060 [0.049, 0.071]				
	6-10 years (n=293)	82.86 (75) ns	1.10	0.995	0.996	0.019 [0.000, 0.040]				
	11+ years (n=413)	120.83 (75)*	1.61	0.975	0.982	0.039 [0.025, 0.051]				
	M1: Configural	390.21 (225)*	1.73	0.970	0.978	0.026 [0.021, .030]	—	—	—	—
	M2: Metric	416.23 (245)*	1.70	0.971	0.978	0.025 [0.021, 0.029]	M1	-0.001	0.000	-0.001 [-0.001, 0.000]
	M3: Scalar	501.41(275)*	1.82	0.966	0.970	0.027 [0.023, 0.031]	M2	0.005	0.008	0.008 [-0.002, -0.002]
M4: Residual	587.91 (299)*	1.97	0.960	0.962	0.029 [0.026, 0.032]	M3	0.006	0.008	-0.002 [-0.002, 0.001]	

Table 7. Continued

Work Tasks Motivation Scale for Teachers (WTMST)- Student Evaluation Tasks										
Gender										
	Male (n=835)	161.86 (75)*	2.13	0.987	0.990	0.037 [0.029, 0.045]				
	Female (n=282)	109.04 (75)*	1.43	0.985	0.989	0.039 [0.021, 0.055]				
MI	M1: Configural	270.99 (152)*	1.78	0.986	0.990	0.026 [0.021, 0.032]	—	—	—	—
	M2: Metric	281.81 (162)*	1.74	0.987	0.990	0.026 [0.021, 0.031]	M1	-0.001	0.000	0.000 [0.000, 0.001]
	M3: Scalar	290.82 (177)*	1.64	0.989	0.990	0.024 [0.019, 0.029]	M2	-0.002	0.000	0.002 [0.002, 0.002]
	M4: Residual	333.49 (192)*	1.74	0.985	0.987	0.026 [0.021, 0.030]	M3	0.004	0.003	0.002 [-0.002, -0.001]
Age										
	25 -35 years (n=695)	179.28 (75)*	2.36	0.984	0.988	0.044 [0.036, 0.053]				
	36 -45 years (n=365)	152.28 (75)*	2.01	0.963	0.973	0.053 [0.040, 0.065]				
	46+ years(n=57)	89.79 (75) ns	1.20	0.948	0.963	0.59 [0.000, 0.091]				
MI	M1: Configural	425.18 (228)*	1.86	0.977	0.984	0.028 [0.024, 0.032]	—	—	—	—
	M2: Metric	464.81 (248)*	1.87	0.977	0.982	0.028 [0.024, 0.032]	M1	0.000	0.002	0.000 [0.000, 0.000]
	M3: Scalar	773.86 (269)*	2.88	0.951	0.958	0.041 [0.038, 0.044]	M2	0.026	0.024	-0.013 [-0.014, -0.012]
	M4: Residual	1003.27(296)*	3.39	0.938	0.941	0.046 [0.043, 0.049]	M3	0.013	0.017	-0.005 [-0.005, -0.005]
University Type										
	Research University (n=431)	135.65 (75)*	1.81	0.983	0.988	0.043 [0.031, 0.055]				
	Applied University (n=333)	140.69 (75)*	1.88	0.972	0.980	0.051 [0.038, 0.064]				
MI	General University (n=353)	138.98 (75)*	1.83	0.979	0.985	0.049 [0.036, 0.061]				
	M1: Configural	408.97 (228)*	1.79	0.980	0.985	0.027 [0.022, 0.031]	—	—	—	—
	M2: Metric	456.97 (245)*	1.87	0.978	0.983	0.028 [0.024, 0.032]	M1	0.002	0.002	-0.001 [-0.002, -0.001]
	M3: Scalar	610.89 (275)*	2.22	0.969	0.973	0.033 [0.030, 0.037]	M2	0.009	0.010	-0.005 [-0.006, -0.004]
	M4: Residual	826.67 (299)*	2.76	0.955	0.957	0.040 [0.037, 0.043]	M3	0.014	0.016	-0.007 [-0.007, -0.005]
Experience in teaching										
	Below 5 years (n=411)	103.88 (75)*	1.37	0.991	0.994	0.030 [0.013, 0.043]	—	—	—	—
	6 - 10 years (n=293)	120.83 (75)*	1.59	0.981	0.986	0.045 [0.029, 0.060]	—	—	—	—
	11+ years (n=413)	169.44 (75)*	2.23	0.972	0.980	0.055 [0.044, 0.065]	—	—	—	—
MI	M1: Configural	394.16 (228)*	1.73	0.981	0.986	0.026 [0.021, 0.030]	—	—	—	—
	M2: Metric	438.87 (248)*	1.77	0.980	0.984	0.026 [0.022, 0.030]	M1	0.001	0.002	0.000 [-0.001, 0.000]
	M3: Scalar	500.97 (275)*	1.82	0.979	0.981	0.027 [0.023, 0.031]	M2	0.001	0.003	-0.001 [-0.001, -0.001]
	M4: Residual	624.33 (305)*	2.05	0.973	0.974	0.031 [0.027, 0.034]	M3	0.006	0.007	-0.004 [-0.004, -0.003]

Note. $N = 1,117$; * $p < 0.001$; MI = measurement invariance, RMSEA = root mean squared error of approximation, TLI = Tucker-Lewis index; CFI = comparative fit index; $\Delta RMSEA$ = change root mean squared error of approximation, ΔTKI = change Tucker-Lewis index; ΔCFI = change comparative fit index.

The Two-Factor Model of WTMST for both Teaching and Student Evaluation Tasks

The hybrid WTMST model of teaching and student evaluation tasks aim is both the teaching and student evaluation tasks are potential positive sources of work task motivation. According to Fernet et al. (2008), teaching tasks and student evaluation tasks of WTMST have a positive correlation. The two-hybrid factor model (see Figure 3) was fitted with this data and the result was as follows: $\chi^2(394) = 1020.49$, $\chi^2/df = 2.59$, $GFI = 0.941$, $TLI = 0.965$, $CFI = 0.968$, $RMSEA = 0.038$, 95% CI [0.035, 0.041]. However, in this study, the two-hybrid models, even though the model fit was good, had a negative correlation and did not support hypothesis 5. The teaching and student evaluation tasks are highly associated and support each other. The result might be due to cultural differences and teachers' motivation preferences in their job.

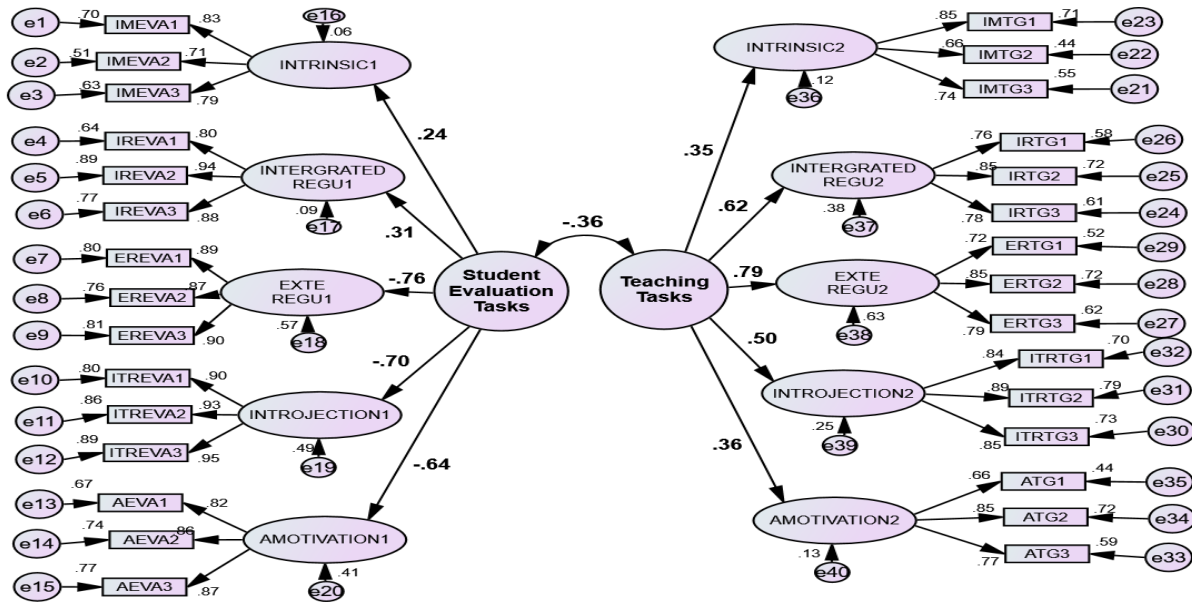


Figure 3. Two Hybrid Factor Model of WTMST.

Discussion

This study made an effort to validate and evaluate the four competitive models (single, correlated, bi-factor, and higher-order) of the WTMST for teaching and student evaluation tasks by developing and examining the culturally-adapted Amharic version of the WTMST with a large sample of instructors from an Ethiopian public university. In contrast to Fernet et al.'s (2008), who investigated a good model fit of WTMST using the correlated factor model, we included the single factor, bi-factor and the higher order models in the diverse sample of public universities. The newly-added and expanded models gave additional insight into WTMST for teaching and student evaluation tasks of its multi-dimensionality nature.

Due to the lack of teacher-specific motivation measures in the Ethiopian context based on the SDT of motivation, and after the reconsideration of the previously-conducted research, the importance and usefulness of the cross-cultural adaptation and validation of an instrument is evident.

The Amharic version of the WTMST developed strictly following the global guidelines for cultural adaptation (Davidov et al., 2014) confirmed high reliability and best construct validity in the three independent competitive models (bi-factor, correlated factor and higher-order) models and proved good convergent validity. In addition, the CFA model of WTMST for both teaching and student evaluation tasks proved that the bi-factor model confirmed a much better fit than the correlated and the higher-order factor models.

Correlated and higher-order factor models were relatively congruent and illustrated the best model fit than the single-factor model. However, due to the unfit of the global cut-off points, the single-factor model was rejected and revealed the worst model fit than the other three CFA competitive models; this advocates WTMST multi-dimensional model. The best fit of the multi-dimensional model of WTMST and the poor fit of the single factor model in teaching and student evaluation tasks indicates the distinct domains' existence. Presumably, the WTMST for teaching and student evaluation tasks is the first study to confirm the multi-dimensionality of the WTMST by evaluating the four different CFA competitive fit models. Furthermore, this study tested for MI across various groups (genders, age, university type and experience in teaching).

The current study used single and multi-group CFA to test the WTMST of teaching and student evaluation tasks that the bifactor model is invariant across gender, university type, and experience in teaching for teaching tasks and student

evaluation. In the first step, single group CFA analyses were conducted and uncovered acceptable fit indices for subgroups of gender (male versus female), age (25–35, 36–45, and 46+ years old), university type (research, applied, and general), and experience in teaching (below 5, 6–10 years and 11+ years). The only exception was the age 46 and above subgroup (n=57), which was slightly lower than the other age sub-group. The measurement invariance was performed following the single group and the multi-group CFA. The models were invariant except for age in teaching and student evaluation tasks of WTMST. The findings of present study regarding MI support the existing literature (Cheung & Rensvold, 2002; Millsap, 2011; Putnick & Bornstein, 2016; Vandenberg & Lance, 2000; van de Schoot et al (2012)..

Conclusion

Besides acknowledging that the adapted Amharic version of WTMST in teaching and student evaluation tasks is psychometrically sound, reliable, valid, and invariant using multi-modal CFA comparisons, this study also offers a sophisticated validated scale exclusively intended for assessing university instructors' motivation to the scientific community. However, it was conducted by the university instructors by taking large-scale data from Ethiopian higher education. This study was also triangulated using various recommended advanced statistical methods to ensure the psychometric properties of WTMST separately and combined both teaching and student evaluation tasks.

Recommendations

The WTMST has to be substantially validated in large-scale data by taking the different types of universities, such as research, applied and general universities. In addition, comparing CFA models in teaching and student evaluation tasks and testing invariants across various groups may also offer solid methodological evidence applicable to educational settings. However, this study targeted only two teacher tasks (teaching and student evaluation) among the six: administrative, classroom preparation, classroom management, and complementary tasks. Because of that, further studies should be advised that focus on the rest of the teachers' tasks based on multi-dimensional WTMST across various languages and cultures. In addition, in this study primary and secondary school teachers are not involved thus future inquiries should also consider this gap and compare the schools to intervene in teachers' motivation. Most notably, in this study, university teachers were required to make the validation and measure invariance in Ethiopian higher educational settings. Therefore, future studies should consider school administrators, students, and teachers to assess the WTMST in teaching and student evaluation tasks to make the best decisions in educational settings.

Furthermore, future studies could extend the present research by incorporating other aspects of teachers' motivational tasks and their association with teacher well-being and stress. Last but not least, the collected data of this study were self-reported using a quantitative approach. Hence, future studies are expected to include both quantitative and qualitative approaches using case studies, interviews, and self-reported measures to triangulate and get context-based motivational problems of teachers to be more valuable.

Limitations

Several drawbacks of the study must be acknowledged. First, this study assessed university teachers only in Ethiopian higher educational settings. However, in this study, we have taken on a global and specific context-based scientific literature dealing with WTMST and the relevant tasks of teachers. Therefore, the WTMST measure should be relevant to explore the association between teacher well-being and burnout. Second, replication of WTMST is needed by considering two or more tasks in various nations would also boost the relevance of the instrument in teaching contexts. Third, this WTMST Amharic version did not yet comprise the four pillars among the six teachers' tasks, i.e., administrative, classroom preparation, classroom management, and complementary tasks of Fernet et al.'s (2008) dimensions. Future studies could fill the gaps of this study by looking for the rest of the teachers' tasks and associating them with teacher well-being in other educational and organizational settings.

Authorship Contribution Statement

Zewude: Conceptualization, data collection and acquisition, statistical analysis and interpretation, writing and editing, critical revision of the manuscript. Hercz: Conceptualization, writing, reviewing and editing, technical and material support, supervision. Duong: Conceptualization, writing, reviewing and editing, technical and material support. Pozsonyi: Conceptualization, writing, reviewing and editing.

References

- Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, 42(5), 815–824. <https://doi.org/10.1016/j.paid.2006.09.018>
- Byrne, B. M. (2010). *Structural Equation Modeling with AMOS: Basic concepts, applications, and programming*. Taylor and Francis Group, LLC.
- Byrne, B. M., & van de Vijver, F. J. R. (2010). Testing for measurement and structural equivalence in large-scale cross-

- cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing*, 10(2), 107–132. <https://doi.org/10.1080/15305051003637306>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bi-factor and second-order models of quality of life. *Multivariate Behavioral Research*, 41(2), 189–225. https://doi.org/10.1207/s15327906mbr4102_5
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233–255. <https://doi.org/fd4n2m>
- Collie, R. J. (2014). *Understanding teacher well-being and motivation: measurement, theory, and change over time*. The University of British Columbia. <https://doi.org/10.1037/t58623-000>
- Collie, R.J., Shapka, J.D., Perry, N.E., & Martin, A. (2015). Teacher well-being: exploring its components and a practice-oriented scale. *Journal of Psychoeducational Assessment*, 33(8), 744–756. <https://doi.org/10.1177/0734282915587990>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>
- Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, 40(2), 55–75. <https://doi.org/gd7fzk>
- Davidov, E., Schmidt, P., Billiet, J., & Meuleman, B. (2018). *Cross-cultural analysis methods and applications*. Taylor & Francis Group. <https://doi.org/10.4324/9781315537078>
- Deci, E. L., & Ryan, R. M. (2008a). Facilitating optimal motivation and psychological well-being across life's domains. *Canadian Psychology*, 49(1), 14–23. <https://doi.org/10.1037/0708-5591.49.1.14>
- Deci, E. L., & Ryan, R. M. (2008b). Self-determination theory: A macro-theory of human motivation, development, and health. *Canadian Psychology*, 49(3), 182–185. <https://doi.org/10.1037/a0012801>
- Eremenco, S. L., Cella, D., & Arnold, B. J. (2005). A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Evaluation and the Health Professions*, 28(2), 212–232. <https://doi.org/10.1177/0163278705275342>
- Fernet, C. (2011). Development and validation of the work role motivation scale for school principals (WRMS-SP). *Educational Administration Quarterly*, 47(2), 307–331. <https://doi.org/10.1177/0013161X10385108>
- Fernet, C., Sencal, C., Guay, F., Marsh, H., & Dowson, M. (2008). The work tasks motivation scale for teachers (WTMST). *Journal of Career Assessment*, 16(2), 256–279. <https://doi.org/10.1177/1069072707305764>
- Gagné, M., & Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior*, 26(4), 331–362. <https://doi.org/10.1002/job.322>
- Gagné, M., Forest, J., Gilbert, M. H., Aubé, C., Morin, E., & Malorni, A. (2010). The motivation at work scale: Validation evidence in two languages. *Educational and Psychological Measurement*, 70(4), 628–646. <https://doi.org/10.1177/0013164409355698>
- George, D., & Mallery, P. (2020). *IBM SPSS statistics 26 step by step: A simple guide and reference*. Routledge. <https://doi.org/10.4324/9780429056765>
- Hair, J., Black, W., Babin, B., & Anderson, R. (2019). *Multivariate data analysis*. Annabel Ainscow.
- Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*. *Electronic Journal of Business Research Methods*, 6(1), 53–59. <https://doi.org/10.1037/1082-989X.12.1.58>
- Hu, L. T., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Immekus, J. C., & Imbrie, P. K. (2008). Dimensionality assessment using the full-information item bifactor analysis for graded response data: An illustration with the state metacognitive inventory. *Educational and Psychological Measurement*, 68(4), 695–709. <https://doi.org/10.1177/0013164407313366>
- Johnson, R. E., Rosen, C. C., & Djurdjevic, E. (2011). Assessing the impact of common method variance on higher order multi-dimensional constructs. *Journal of Applied Psychology*, 96(4), 744–761. <https://doi.org/10.1037/a0021504>
- Kim, H.-Y. (2013). Statistical notes for clinical researchers: assessing normal distribution using skewness and kurtosis. *Restorative Dentistry & Endodontics*, 38(1), 52. <https://doi.org/10.5395/rde.2013.38.1.52>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling*. The Guilford Press.

- Lei, P.-W., & Wu, Q. (2007). Introduction to structural equation modeling: Issues and practical considerations. *Educational Measurement: Issues and Practice*, 26(3), 33–43. <https://doi.org/10.1111/j.1745-3992.2007.00099.x>
- Liang, X., & Luo, Y. (2020). A Comprehensive comparison of model selection methods for testing factorial invariance. *Structural Equation Modeling*, 27(3), 380–395. <https://doi.org/10.1080/10705511.2019.1649983>
- Maslach, C., Jackson, S. E., & Leiter, M. (1996). *Maslach burnout inventory: Educators survey*. Palo Alto, CA: Consulting Psychologists Press.
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Routledge. <https://doi.org/10.4324/9780203821961>
- Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of Cardiac Anaesthesia*, 22(1), 67–72. https://doi.org/10.4103/aca.ACA_157_18
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41(2), 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1), 54–67. <https://doi.org/10.1006/ceps.1999.1020>
- Ryan, R. M., & Deci, E. L. (2017). *Self-Determination Theory: Basic psychological needs in motivation, development, and wellness*. The Guilford Press. <https://doi.org/10.1521/978.14625/28806>
- Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia and Analgesia*, 126(5), 1763–1768. <https://doi.org/gfbxr2>
- Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, 42(5), 893–898. <https://doi.org/10.1016/j.paid.2006.09.017>
- Stockdale, G. D., Gridley, B. E., Balogh, D. W., & Holtgraves, T. (2002). Confirmatory factor analysis of single- and multiple-factor competing models of the dissociative experiences scale in a nonclinical sample. *Assessment*, 9(1), 94–106. <https://doi.org/10.1177/1073191102009001011>
- van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, 9(4), 486–492. <https://doi.org/10.1080/17405629.2012.686740>
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4–70. <https://doi.org/10.1177/109442810031002>
- Wammerl, M., Jaunig, J., Mairunteregger, T., & Streit, P. (2019). The German version of the PERMA-profiler: Evidence for construct and convergent validity of the PERMA theory of well-being in German-speaking countries. *Journal of Well-Being Assessment*, 3(2–3), 75–96. <https://doi.org/10.1007/s41543-019-00021-0>
- Wang, Y., Kim, E. S., Dedrick, R. F., Ferron, J. M., & Tan, T. (2018). A multilevel bi-factor approach to construct validation of mixed-format scales. *Educational and Psychological Measurement*, 78(2), 253–271. <https://doi.org/10.1177/0013164417690858>
- Zewude, G. T., & Hercz, M. (2021). Psychological capital and teacher well-being: The mediation role of coping with stress. *European Journal of Educational Research*, 10(3), 1227–1245. <https://doi.org/10.12973/eu-jer.10.3.1227>
- Zewude, G. T., & Hercz, M. (2022). The teacher well-being scale (TWBS): Construct validity, model comparisons and measurement invariance in an Ethiopian setting. *Journal of Psychology in Africa*, 32(3), 251–262. <https://doi.org/jbw5>

Appendix A

Amharic Version (የመምህራን የሥራ ተነሳሽነት መለኪያ; Work Task Motivation Scale for Teachers)

መሠረታዊ፦ ቀጥሎ በተመለከተው ሰንጠረዥ ውስጥ የእርስዎን የሥራ ተነሳሽነት የሚገለጹ ዓ.ነገር ተዘርዘሯል። በእያንዳንዱ ዓ.ነገር ፊትለፊት ድጋፍን የመፈጸም ሁኔታ የሚያመለክቱ (ከ1-7) ነጥቦች) ሚዛ ተቀምጧል። ከ1-7 ከቀረቡት ነጥቦች ያንችን/ተን ሁኔታ በትክክል ይገልጻል ብለህ/ሽ የምታስበውን/ዊን ነጥብ እያንዳንዱ ዓ.ነገር ፊት ለፊት ከተቀመጡት አማራጮች አንዱን ክበብ/ቢ። የሚከተሉትን ሚዛኖች የሚወክሉት 1 = በፍጹም አይደሉም ፣ 2 = በጣም ትንሽ ይዘመዳል ፣ 3 = ትንሽ ይዘመዳል ፣ 4 በመጠኑ ይዘመዳል ፣ 5 = በጣም ትንሽ ይዘመዳል ፣ 6 = በጣም በጣም ትንሽ ይዘመዳል ፣ 7 = ሙሉ በሙሉ ይዘመዳል።

No.	Teaching task (የማስተማር ተግባር)	Student evaluation task (የተማሪ ግምገማ ተግባር)
	ከሚከተሉት ምክንያቶች ውስጥ እያንዳንዳቸው ምን ያህል እንደሚመሰሰሉ ያሳያል። ለምን የማስተማር ሙያ ይሠራሉ?	
1	ምክንያቱም ማስተማር ደስ የሚል ሙያ ስለሆነ ነው።	ምክንያቱም ይህን ተግባር መፈጸም ደስ የሚል ነው።
2	ምክንያቱም ማስተማር አስደሳች ሆኖ ስለገኘሁት ነው።	ምክንያቱም ይህን ሥራ መሥራት አስደሳች ሆኖ አግኝቼዋለሁ
3	ምክንያቱም የማስተማር ሙያ ስለምወድ ነው።	ምክንያቱም ይህን ተግባር መሥራት እወዳለሁ።
4	ምክንያቱም የማስተማር ስራ መስራት ለእኔ አስፈላጊ ስለሆነ ነው።	ምክንያቱም ይህን ተግባር መወጣት ለእኔ አስፈላጊ ነው
5	ምክንያቱም የማስተማር ስራ አስፈላጊ ነው ብዬ የማስበውን የሥራዬን ዓላማ እንዳሳካ ስለሚያስችለኝ ነው።	ምክንያቱም ይህ ተግባር አስፈላጊ ብዬ የማስበውን የሥራ ዓላማ እንዳሳካ ስለሚያስችል ነው።
6	ምክንያቱም ይህ ተግባር ለተማሪዎቼ ትምህርታዊ ስኬት አስፈላጊ ሆኖ አግኝቼዋለሁ።	ምክንያቱም ይህ ተግባር ለተማሪዎቼ አካዴሚያዊ ስኬት አስፈላጊ ሆኖ አግኝቼዋለሁ
7	ምክንያቱም ይህን ተግባር ካላከናወንኩ መጥፎ ስሜት ይሰማኛል።	ምክንያቱም ይህን ተግባር ካላከናወንኩ መጥፎ ስሜት ይሰማኛል
8	ምክንያቱም ባለማድረግ የጥፋተኝነት ስሜት ይሰማኛል።	ምክንያቱም ባለማድረግ የጥፋተኝነት ስሜት ይሰማኛል።
9	ካለስተማርኩ መጥፎ ስሜት እንዳይሰማኝ ነው።	ካለደረግኩት መጥፎ ስሜት እንዳይሰማኝ ነው።
10	ምክንያቱም ስራዬ ይፈልገኛል።	ምክንያቱም ሥራዬ ይፈልገዋል
11	ምክንያቱም ትምህርት ቤቱ እንድሠራ ያስገድደኛል።	ምክንያቱም ትምህርት ቤቱ እንድሠራ ያስገድደኛል።
12	ምክንያቱም እኔ ለምሠራው ሥራ ስለ ሚከፈለኝ ነው።	ምክንያቱም እኔ ለማድረግ የሚከፈልበት ነው።
13	አላውቅም ብቻ፡ የማስተማር ስራ አስፈላጊነት ሁልጊዜ አይታዩኝም።	አላውቅም, ይህንን ተግባር የመፈጸም አስፈላጊነት ሁልጊዜ አይታዩኝም
14	የማስተማር ስራ ለምን እንደሰራሁ አውቃለሁ ፣ ምክንያቱን ከአሁን በኋላ ግን አይታዩኝም።	ይህንን ተግባር ለምን እንደሰራሁ አውቃለሁ ፣ ግን ምክንያቱን ከአሁን በኋላ አይታዩኝም።
15	አላውቅም፡ አንዳንድ ጊዜ የማስተማር ስራ ዓላማው አይታዩኝም።	አላውቅም, አንዳንድ ጊዜ ተማሪዎችን መመዘን ዓላማው አይታዩኝም።

Appendix B

English Version (Work Task Motivation Scale for Teachers)

Direction: Using the scale below indicates to what extent each of the following items presently corresponds to one of the reasons, Why you are doing the teaching and ? And then, encircle one of the Seven alternative numbers that best describes Does not correspond at all or Correspond completely with the statement about yourself.

The numbers represent 1 = It does not correspond at all, 2 = Correspond very little, 3 = Correspond a little, 4 = Correspond moderately, 5 = Correspond strongly, 6 = Correspond very strongly, 7 = Correspond completely

No.	Teaching task	Student evaluation task
1	Because it is pleasant to carry out this task	Because it gives me happiness to carry out this task
2	Because I find this task interesting to do	Because I find this task interesting to do
3	Because I like doing this task	Because I like doing this task
4	Because it is crucial for me to carry out this task	Because it is important for me to carry out this task
5	Because this task allows me to attain work objectives that I consider important	Because this task allows me to attain work objectives that I consider important.
6	Because I find this task important for the academic success of my students	Because I find this task important for the academic success of my students
7	Because if I do not carry out this task, I will feel bad	Because if I don't carry out this task, I will feel bad
8	Because I would feel guilty not doing it	Because I would feel guilty not doing it
9	To not feel bad if I do not do it	To not feel bad if I don't do it
10	Because my work demands it	Because my work demands it
11	Because the school obliges me to do it	Because the school obliges me to do it
12	Because I'm paid to do it	Because I'm paid to do it
13	I do not know; I do not always see the relevance of carrying out this task	I don't know, I don't always see the relevance of carrying out this task
14	I used to know why I was doing this task, but I don't see the reason anymore	I used to know why I was doing this task, but I don't see the reason anymore
15	I do not know; sometimes, I do not see its purpose	I don't know, sometimes I don't see its purpose

Note: Intrinsic motivation = 1, 2, 3; identified regulation = 4, 5, 6; introjected regulation = 7, 8, 9; external regulation = 10, 11, 12; amotivation = 13, 14, 15

Source: Fernet, C., Senécal, C., Guay, F., Marsh, H., & Dowson, M. (2008). The work tasks motivation scale for teachers (WTMST). *Journal of Career Assessment*, 16(2), 256–276. <https://doi.org/10.1177/1069072707305764>