

Psychometric Properties and measurement invariance by gender of the Teacher Competency Evaluation Scale for Physical Education in the Mexican Context

Propiedades psicométricas y medición de la invarianza por género de la Escala de Evaluación de Competencias Docentes de Educación Física en el contexto mexicano

*Raúl Baños, **Juan Pablo Machado-Parra, ***Juan Carlos Pérez-Morán, **Michelle Barretos-Ruvalcaba, ****Antonio Baena-Extremera

* Universidad de Zaragoza (España), ** Universidad Autónoma de Baja California (México), *** Red Impulsora de Metodología en Evaluación Diagnóstica e Innovación Educativa (México), ****Universidad de Granada (España)

Resumen. El presente es un estudio de obtención de evidencias de validez, cuyo propósito fue analizar las propiedades psicométricas y medir la invarianza según el género de la Escala de Evaluación de Competencias Docentes de Educación Física (EECD-EF) para escuelas secundarias en el contexto mexicano. Los participantes fueron 748 estudiantes de secundaria con una edad media de 13.99 años, de los cuales 374 eran mujeres. Los resultados del Análisis Factorial Confirmatorio (AFC) apoyan el modelo unifactorial. La EECD-EF cumple los requisitos de validez convergente. Además, la correlación con la subescala de *Satisfacción de la Escala de Satisfacción Intrínseca en la Escuela* (ISDS, por sus siglas en inglés) fue positiva y significativa, proporcionando evidencia de validez de asociación con otras variables. Sin embargo, no se encontró una correlación discriminante con la subescala de *Aburrimiento*. Se concluye que la EECD-EF sirve como instrumento eficaz para evaluar la invarianza de la medida de la competencia de los profesores desde el punto de vista de los alumnos en función del género.

Palabras clave: Competencia docente, Educación física, Validez, Satisfacción intrínseca, Análisis Factorial Confirmatorio, Invarianza factorial.

Abstract. This is a study to obtain evidence of validity, whose purpose was to analyze the psychometric properties and the measurement invariance by gender of the *Teacher Competency Evaluation Scale for Physical Education* (EECD-EF, by its acronym in Spanish) for secondary schools in the Mexican context. Participants were 748 high school students with a mean age of 13.99 years, of whom 374 were female. The results from the Confirmatory Factor Analysis (CFA) support the unifactorial model. The EECD-EF meets the requirements for convergent validity. In addition, the correlation with the *Satisfaction* subscale of *Intrinsic School Satisfaction Scale* (ISDS) was positive and significant, providing evidence of validity of association with other variables. However, no discriminatory correlation was found with the *Boredom* subscale. It's concluded that the EECD-EF serves as an efficient instrument to assess the measurement invariance of the teachers' competency from the students' point of view according to gender.

Keywords: Teacher competency, Physical education, Validity, Intrinsic Satisfaction, Confirmatory Factor Analysis, Factorial invariance.

Introduction

Competency-based teaching involves new approaches for the teaching-learning process on both the curricular and the methodological innovation levels (Baena-Extremera, Granero-Gallegos & Martínez-Molina, 2015). Moreover, analyzing the teacher's role becomes more relevant due to the high expectations from education systems with regard to teachers acquiring competencies that allow them to address and solve issues related to changes resulting from the learning process through time and from professional improvement, with the possible reconstruction of reality and an effective personal and professional positioning (Palacios, López-Pastor & Fraile, 2019). This is why personal and professional traits, and the teachers' actions and behaviors are relevant topics for comprehensive study and surveying

when facing the novel education scenarios (Baños et al., 2017; Barba-Martín, et al., 2020).

For the 21st century education model, the Organization for Economic Co-operation and Development (OECD) deemed relevant to include the Teaching and Learning Survey project (TALIS), the importance of developing more efficient teaching competencies (OECD, 2013; Ainley & Carstens, 2018). Such efficient competencies have to do with the "specific attributes of teaching performance related to functions and responsibilities inherent to everyday tasks (the generic), but, most of all, tightly articulated to a specific teaching practice, that is, linked to the specific and authentic tasks of a unique teaching practice" (Blázquez, 2013, p. 4). Authors such as Rivera and Alfageme-González (2019) consider that teaching competencies are the result of the professional training background, but also of learnings acquired through life, at school or elsewhere, during either the teaching tasks themselves or one's

leisure time. In this manner, according to Tejada (2005), a competent teacher possesses four know-hows: (1) the theoretical *know-how* regarding an academic field within the professional framework; (2) the *know-how to apply* said knowledge to specific professional situations; (3) the *know-how to be* from the personal attributes and attitudes toward oneself, the students and the teaching practice; (4) the *know-how to act* regarding the cumulus of interpersonal demeanor and skills that allow them to interact in the professional environment.

One method to assess teaching competencies can be seen in the yearly teacher selection processes in Mexico. In order to be appointed for a teaching position at elementary schools, the *Secretariat of Public Education* (SEP, 2017) uses a knowledge and skills test in which applicants must prove that they meet the profile, parameters and indicators set out by the applicable normativity. This seeks to promote the best learning achievements for all students.

The teaching profile established by the Mexican normativity comprises five dimensions: (1) a teacher that knows their students, how they learn and what they need to learn; (2) a teacher that organizes and evaluates learning tasks and performs an appropriate didactic approach; (3) a teacher self-identifying as a professional in constant improvement to help students during their learning processes; (4) a teacher that abides by the legal and ethical commitments inherent to their profession for the wellbeing of their students; and (5) a teacher that participates in the efficient operation of the school and fosters its relationship with the community to ensure that all students can graduate successfully. These capabilities should be present in the everyday tasks of Mexican teachers, however, many of these attributes have either faded over time or never been reinforced. This is why it is important to be continuously aware of the current teaching competency levels (Baños, Barretos-Ruvalcaba & Baena-Extremera, 2019).

Students are an essential factor for the preservation and improvement of quality teaching due to their participation in evaluation processes that guarantee the internal quality of teaching (Lidice & Saglam, 2012). However, most existing assessments are based on a traditional approach where students evaluate the learnings received throughout the course by term's end (Marsh et al., 2011). Hence, researchers have become increasingly worried about the lack of reliability and validity of some potentially biased questionnaires based on the students' point of view (Catano & Harvey, 2011).

For several years now, scientific literature has strived to assess competencies and standards for teachers (OXFAM NOVID, 2011), creating models such as the *Competency Based Teacher Education* (Navío, 2005) with the goal of determining what kind of competencies will be present when

training teachers with a general profile (Catano & Harvey, 2011; Luna et al., 2012) and teachers with a specific profile in Physical Education (PE) (Gallego-Ortega & Rodríguez-Fuentes, 2017; Lleixá et al. 2010). Likewise, several assessment instruments have been created to measure the competencies of PE teachers (Ojeda-Nahuelcura et al., 2022), obtaining acceptable values for validity and reliability of some questionnaires and scales. For example, in the studies of Guillén-Gámez and Perrino (2020), and Perea and Abello (2022) measured digital skills, Tul et al. (2015) focused on professional competences, and Jones-Jofré et al. (2022) in generic ones. For his part, Martínez-Mínguez et al. (2022) analyze for the scores of an instrument to assess the professional psychomotor skills in the context of the initial teacher training.

Specifically, in Chile, Jones-Jofré et al. (2022) applied in 2017 the *Physical Education Generic Competencies Questionnaire* (PEGCQ) to 961 first-year students belonging to the different careers of the Technical University Federico Santa María de Valparaíso, to know the effect of physical education classes on the development of generic competences through the perception of students. The competencies evaluated were: (a) *Teamwork*, (b) *Recreation*, (c) *Kinesthetic intelligence* and (d) *Ethics*. Among the conclusions, it is highlighted that users affirm that physical education classes allow the development of these generic skills through the systematic practice of physical activity, thus contributing to the integral formation of the individual. Similarly, Tul et al. (2015) designed a cross-sectional study to evaluate perceived importance of Slovenian Physical Education Teachers' (PETs) professional competencies. In a self-administered questionnaire 672 Slovenian PETs evaluated their own professional competencies on a four-level scale. The results point to a high degree of reliability of the entire questionnaire (Cronbach alpha=.97). What's more, the competency profile of teachers consists of 11 factors: (a) *Didactic approaches*; (b) *Social science aspects of sport*; (c) *General pedagogical knowledge*; (d) *Biological and physiological aspects of sport*; (e) *Teaching methods*; (f) *General didactic knowledge*; (g) *Research, entrepreneurship, organization*; (h) *Communication skills*; (i) *Planning*; (j) *Responsible behavior*; and (k) *Leadership, motivation*. *Didactic approaches* as the subject-specific competencies, which represent the first factor, seemed to be the most informative for PETs, explaining 33.8% of the total variance (57.6%).

In Spain, Martínez-Mínguez et al. (2022) obtained evidence of validity and reliability for the scores of *Professional Psychomotor Competencies in Initial Teacher Training Scale* (PPCITTS), applied to 318 early childhood education teacher students from four Spanish universities. The results of the validity analysis of the internal structure yielded a distribution in five interpretable factors capable of jointly

explaining 67.4% of the total variance, which are: (a) *Design psychometricity sessions*, (b) *Apply the adult attitude system*, (c) *Observe and evaluate the child and the session*, (d) *Analyze and reflect on their own practice and that of others*, and (e) *Interact with the educational and scientific community*. Regarding the reliability analysis, the instrument presents high coefficients of internal consistency by scales and evidence of temporal stability.

Among the instruments devised, we highlight the *Assessment Questionnaire for Teaching Performance* (CEID, by its acronym in Spanish) with a factorial structure of three dimensions: planning, development and results (Moreno-Murcia et al., 2015), which yielded acceptable validity and reliability values. On the other hand, Salcines-Talledo et al. (2018) created the *Initial Training of Physical Education Teachers Scale* (ITPES) without reporting the goodness-of-fit indices of the Confirmatory Factor Analysis (CFA), obtaining three dimensions: (a) *Cross-competencies*, (b) *Teaching competencies*, and (c) *PE teaching competencies*. Moreover, the *Interpersonal Competency Valuation Scale* (ICVS) for PETs was created featuring four factors: (1) teamwork, (2) communication with experts from other areas, (3) skills for interpersonal relations, (4) appreciation of diversity and multiculturalism, (5) auto-critical and critical capacity, and (6) ethical commitment (Aparicio & Fraile, 2015). In this scale, neither validity nor reliability results were reported for the instrument. All participants were university students.

PE teaching evaluation instruments are scarce at a secondary school level, with the only highlight being the Spanish version of the *Teacher Competency Assessment Scale for Physical Education* (EECD-EF, by its acronym in Spanish), comprised of eight factors: (a) *Communication*, (b) *Work awareness*, (c) *Creativity*, (d) *Feedback*, (e) *Individual consideration for the student*, (f) *Professionalism*, (g) *Problem solving*, and (h) *Social awareness* (Baena-Extremera et al., 2015). However, in spite of the sound psychometric properties of the EECD-EF applied to the Spanish population and its use in recent studies (Baños, Baena-Extremera & Ortiz-Camacho, 2019; Granero-Gallegos, Baños, Baena-Extremera & Martínez-Molina, 2020), several adaptations are needed for its application and interpretation in the Mexican context. For instance, it is important to highlight the differences not only in student habits, but also in the curricular inception of study programs, as well as in the methodological development of the teaching practice and even the understanding of PE.

Therefore, an adaptation to the Mexican context is needed in order to study the validity and reliability of the EECD-EF and analyze the outlook from Mexican adolescents on their PETs' competencies. In addition, analyzing the measurement factorial invariance delivers strong evidence that allows for a benchmark analysis among popu-

lation subsets according to different interest variables (i.e. *gender*, *cultural background*, and *age*). In particular, the obtained results bring substantial information to continue the analysis of differences in the perception of teaching competencies according to *gender*, given that the emotional and relationship needs between teachers and students are a relevant research topic at the moment (McClowry et al., 2013).

Therefore, the purposes of the present study are varied and respond to different aims and objectives. First, to analyze the psychometric properties of the EECD-EF for Mexican high school students; second, to verify the factorial structure of the unifactorial model underlying the EECD-EF; third, to analyze the factorial invariance in the *configurational*, *weak*, *strong* and *strict* models according to the *gender* of the participants; and the fourth and last, to obtain evidence of relationships with other variables for the EECD-EF.

Method

Type of study

In the present study, evidence of construct validity of the internal structure (internal consistency reliability, dimensionality and factorial invariance) as well as evidence of the relationship with other variables were obtained for the EECD-EF instrument. To obtain evidence of validity of the internal consistency reliability and dimensionality of the test, the recommendations of Brown (2015), Hair et al. (2019), and Hu and Bentler (1999) were followed to define the dimensionality of the underlying model of the instrument. For this purpose, a *Multigroup Confirmatory Factor Analysis* (MGCFA) was applied. Model fit was assessed according to the criteria recommended by Vandenberg and Lance (2000), and Cheung and Rensvold (2002). The factorial invariance of the instrument as a function of gender was assessed according to the practices, recommendations, and criteria recommended by Dimitrov (2010), Milfont and Fischer (2010), and Putnick and Bornstein (2016). To obtain evidence of relationship with other variables, the recommendations of Hair et al. (2019), de Winter et al. (2016), and Liddell and Kruschke (2018) were followed for the selection of the type of correlation test to be used in the simple linear regression analysis.

Participants

The sample was composed of third grade secondary school students from the state of Nuevo Leon (Mexico) chosen by using a probabilistic and multistage design, stratified at the level of school centers and by proportional affixation. According to official data from INEGI (2019), school enrollment in secondary education (ISCED level 2)

for the state of Nuevo León in the 2018-2019 school year was 27,227 students: 13,831 males (50.8%) and 13,396 (49.2%) females. Based on previous data, the sample population was calculated using Cochran's formula (1963) adjusted for finite population correction for proportions. Two groups were considered: men and women; likewise, a confidence level of 95% and a margin of error of +5% were established. As a result, a population of 373 +5 students per group was required.

The sample consisted of 374 male and 374 female third grade high school students from the state of Nuevo Leon (Mexico). Mean age for the male group was 13.99 years (SD = 0.30) and the mean age for the female group was 14.02 years (SD = 0.33).

Instruments

Two instruments were used for the present study: (1) the *Teaching Competence Evaluation Scale in Physical Education for Secondary School* (EECD-EF, by its acronym in Spanish) in its version adapted to the Mexican context (Baena-Extremera et al., 2015) and (2) the *Intrinsic School Satisfaction Scale* (ISDS), in its version adapted for the subject of PE and to the Mexican school context (Baños, Baena-Extremera & Granero-Gallegos, 2019). The EECD-EF is a self-report scale made up of nine items ($k = 9$) that measure the students' perspective on the competencies of physical education, mathematics and Spanish teachers (Baena-Extremera et al., 2015). Each item counts measures the degree of student agreement through seven response options ranging from 1 (*totally disagree*) to 7 (*totally agree*); and, in turn, these are organized into three categories: low agreement (options 1 and 2), medium agreement (options 3 and 4) and high agreement (options 6, 7). Several studies with a Mexican student population have reported adequate evidence of construct validity of the EECD-EF in the Mexican context. Baena-Extremera et al. (2015) report adequate evidence of internal consistency, as they obtained internal consistency indices with adequate values (composite reliability = 0.93; average variance extracted [AVE] = 0.61; Cronbach's alpha [α] = 0.90), as well as adequate fit indices in the single-factor model ($\chi^2 = 27.68$; $gl = 20$; $\chi^2/gl = 1.38$; GFI = 0.99; NFI = 0.98; NNFI = 0.99; CFI = 0.99; RMSEA = 0.023; RMSR = 0.029).

The second instrument used was the ISDS (Baños, Baena-Extremera & Granero-Gallegos, 2019). This instrument is a self-report instrument consisting of eight items ($k = 8$) divided into two factors: (a) a first factor composed of five items that measures the level of *satisfaction/enjoyment* with the academic activities of each subject; and (b) a second factor composed of three items that refers to the students' *boredom* with the academic activities of each subject. The response options were measured through an attitudinal

scale ranging from 1 (*totally disagree*) to 5 (*totally agree*). Evidence of construct validity has been reported in different studies. Baños, Baena-Extremera and Granero-Gallegos (2019) report adequate internal consistency indices ($\alpha > 0.60$, McDonald's Omega [ω] > 0.70), as well as adequate fit indices for the two-factor model ($\chi^2 = 34.30$; $gl = 19$; $\chi^2/gl = 1.80$; GFI = 0.99; NFI = 0.94; NNFI = 0.96; CFI = 0.97 and RMSEA = 0.04).

Procedure

This research was conducted in accordance with the Declaration of Helsinki 1961 (Edinburgh revision, 2008). Approval for this study was obtained from the SEP and the Autonomous University of Baja California (identification number: 431/569/E). To carry out this study, a research project entitled "*Program for International Student Assessment: relationship between school performance in high school students and psychological, family and physical activity variables*" was first submitted and subsequently approved and subsidized by the SEP. Data was collected in September 2019. Then, the authorization of the high school principals was requested, providing the parents/guardians involved with information for consent detailing the purpose and intentionality of the study. After their approval, the data collection procedure was initiated by informing the students that participation was anonymous and voluntary, and that their responses would be confidential, reminding them that there were no right or wrong answers, and asking them to respond with complete honesty. All questionnaires were completed inside the classroom in the presence of the principal investigator in case any doubts arose during the procedure, which lasted between 15 and 20 minutes.

Data analysis

Analyses were performed to obtain descriptive statistics: the mean score, Standard Deviation (SD), Item-Total Correlation Coefficient (ITCC), kurtosis and skewness were calculated for each of the items of the EECD-EF. The assumptions of linearity and multicollinearity, multivariate normality, discrimination, internal consistency and sample adequacy were then verified. Multivariate normality was assessed using Mardia's (1970) multivariate normality test for kurtosis), a non-significant value ($p > 0.05$) indicates that the data are normally distributed (Mardia, 1970). Discrimination was assessed by means of the corrected item-total correlation coefficient (ITCC); those items with values equal to or greater than 0.30 were retained (ITCC ≥ 0.30). The internal consistency assumption was assessed using Cronbach's alpha statistic (α); the criterion for acceptance of the assumption was determined by a value $\alpha \geq 0.70$ (Hair et al., 2019). Finally, the measure of sampling adequacy (MSA) was measured through the Kaiser-Me-

yer-Olkin test (KMO), an MSA value $\geq .70$ and a p-value ≤ 0.50 in Bartlett’s test of sphericity indicates an acceptable degree of association between variables.

Once the assumptions were evaluated, the model underlying the EECD-EF was defined and evaluated. Previous studies (Baños, Baena-Extremera & Ortiz-Camacho, 2019) have obtained adequate fit indices for a unifactorial solution (GFI = 0.99, NFI = 0.96, NNFI = 0.97, CFI = 0.98, RMSEA = 0.003). Therefore, a unifactorial model was generated and its fit was verified. Due to the violation of the multivariate normality assumption and the use of ordinal responses, the weighted least squares (WLS) robust estimation method was used. To measure model fit, the recommendations of Hu and Bentler (1999) for the selection of relative and absolute fit indices were applied. The indices considered were: Satorra-Bentler scaled chi-square test (χ^2/df), normalized fit index (NFI), non-normalized fit index (NNFI), comparative fit index (CFI), root mean square error of approximation (RMSEA), normalized root mean square residual (SRMR). The criteria for verifying adequate fit values for each index were set at: a χ^2/df value equal to or greater than three ($\chi^2/df \leq 3$); NFI, NNFI, and CFI values equal to or greater than 0.95 (NFI, NNFI, CFI ≥ 0.95); and RMSEA and SRMR values equal to or less than .06 (RMSEA, SRMR $\leq .06$). Regarding the CFA of the scales with ordinal nature of the data correlation matrix, it is also important to set out the results of composite reliability and AVE for each of the critical dimensions. According to Hair et al. (2019), the minimum values for composite reliability and AVE should be 0.70 and 0.50 respectively.

Subsequently, it was evaluated whether the construct underlying the test was conceptualized in the same way across participants. For this purpose, a *Multigroup Confirmatory Factor Analysis* (MGCFA) was performed in order to measure the invariance of the construct as a function of the sex of the students based on the one-factor model described above. The recommendations of Milfont and Fischer (2010), Putnick and Bornstein (2016), Dimitrov (2010) were followed. A sequential constraint procedure with increasingly restrictive parameters was followed for the models in each group. Models were nested and compared both models for obtaining *configurational* invariance (M1), i.e., verifying whether the number of factors and patterns of factor loadings are equal. The models were then evaluated for evidence in metric invariance (M2), models invariant in factor loadings; scalar (M3), models invariant in item intercepts; and item uniqueness (M4), models invariant variances/covariances of item errors (Dimitrov, 2010; Putnick & Bornstein, 2016). Criteria for verification of invariance between models were a CFI difference less than -0.01 ($\Delta CFI < -0.01$), an RMSEA di-

ference less than .015 ($\Delta CFI < 0.015$), and a TLI difference less than .01 ($\Delta TLI < 0.01$; Chen, 2007; Putnick & Bornstein, 2016).

Finally, a simple linear regression analysis was carried out to determine the degree of association between the EECD-EF scores and each of the ISDS factors. Spearman’s correlation test (r_s) was applied for this purpose; the application of r_s is recommended for investigating the relationship between two sets of ordinal type data, both normally distributed (de Winter et al., 2016). The correlation between scales with five and seven response options presents results equivalent to the comparison of scales with the same number of response options (Colman et al., 1997), so no transformation of the data was performed to achieve equivalence between scales.

Results

Descriptive analyses of the items

The procedure used to analyze each of the EECD-EF items followed the guidelines set by Carretero-Dios and Pérez (2007). Said analysis is needed to study the convenience of maintaining each item within each of the theoretical dimensions to which it belongs, according to the original scale (Catano & Harvey, 2011).

Values for mean score, SD, ITCC, kurtosis and skewness were calculated for each of the EECD-EF items. Items from the EECD-EF showed mean values ranging from 3.82 (item 4) to 4.3 (item 7), standard deviation values ranged from 1.83 (items 1 and 9) to 1.99 (item 4). As suggested by Bollen and Long (1994), the asymmetry and kurtosis indices are near 0 and $< |2|$; with the exception of the skewness value in item 4 all values showed a negative score trend, indicating negatively skewed and leptokurtic response distributions. All ITCC values of the items were ≥ 0.30 . Complete descriptive statistics for each item are shown in Table 1.

Table 1
Descriptive statistics, internal consistency and homogeneity (n = 748)

Items	M	SD	ITCC	α sin ítem	Skewness	Kurtosis	FL
Ítem 1	4.17	1.83	0.59	0.91	-0.08	-1.03	0.67
Ítem 2	4.17	1.88	0.68	0.91	-0.05	-1.07	0.75
Ítem 3	4.27	1.98	0.73	0.91	-0.26	-1.11	0.79
Ítem 4	3.82	1.99	0.67	0.91	0.15	-1.30	0.74
Ítem 5	4.08	1.96	0.76	0.90	-0.09	-1.19	0.82
Ítem 6	4.12	1.90	0.74	0.91	-0.08	-1.08	0.81
Ítem 7	4.30	1.98	0.74	0.91	-0.23	-1.12	0.81
Ítem 8	4.11	1.87	0.74	0.91	-0.10	-1.09	0.81
Ítem 9	4.17	1.83	0.71	0.91	-0.08	-1.24	0.78

Note: n = sample; M = mean; SD = standard deviation; ITCC = item-total correlation coefficient of; FL = factorial loadings

The mean score for the subjects on the EECD-EF scale was 37.1 with a SD of 13.5. The mean score for fema-

les was 36.9 (SD = 12.8) and for males was 37.3 (SD = 14.1); comparison of mean scores through Student's *t*-test did not yield significant results ($t = -0.44$, $gl = 745.7$, $p = 0.65$).

The value obtained for the normalized multivariate kurtosis of the scale was 28.45 with a significant *p*-value ($p < .05$), this result rejects the assumption of multivariate normality. Given the above, a robust estimation method will be used to generate the underlying model of the test.

The analysis to verify the internal consistency assumption of the test yielded adequate results; the internal consistency was appropriate ($\alpha = 0.92$) and did not improve by removing any item. All ITCC presented values ≥ 0.59 . Factorial saturations of the factor analysis were above 0.65. As well, EECD-EF test yielded significant results for the Bartlett sphericity test ($\chi^2 = 3790.724$, $gl = 36$; $p < 0.001$), whereas Kaiser-Meyer-Olkin was above 0.60 (MSA = 0.94), signaling data adequacy.

Confirmatory Factor Analysis

Results of multivariate normality analysis for the EECD-EF showed that multivariate normality cannot be deemed acceptable, which implies the use of robust estimators. Thereby, we used the *weighted least squares* (WLS) estimation method from the LISREL 8.80 software by Jöreskog and Sorbom (2003). The polychoric correlation and asymptotic covariance matrices were used as input for the data analysis. A single factor measurement model was hypothesized. The tested model exhibits high factorial loads (≥ 0.67). All items show values > 0.50 for individual reliability (R^2) (ranging from 0.50 for item-1 to 0.79 for item-8). All fitness data were outstanding ($\chi^2 = 115.79$, $gl = 27$, $p = 0.000$, $\chi^2/gl = 4.29$, $GFI = 0.99$, $NFI = 0.96$, $NNFI = 0.96$, $CFI = 0.97$, $RMSEA = 0.06$; $RMSR = 0.029$). The model meets all the requirements that guarantee its convergent validity (Hair et al., 2019): all factorial loads were > 0.60 (Figure 1) and statistically significant ($t > 1.96$). Finally, the tested model had a *composite reliability* of 0.93 and AVE of 0.60.

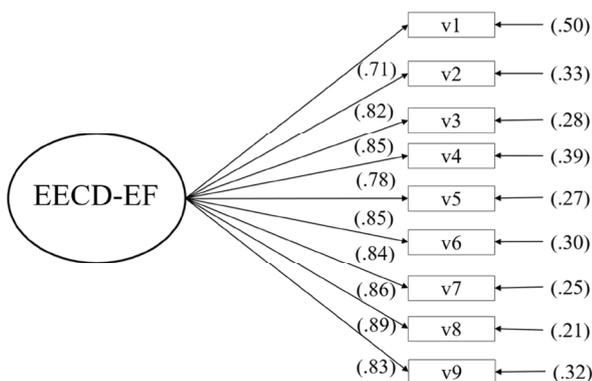


Figure 1. Path Diagram of EECD-EF scale

Factorial invariance according to gender

An MGCFA was performed for the measurement factorial invariance according to *gender* of the EECD-EF. The results from the invariance tests are shown on Table 2. The models (*configural*, *weak*, *strong* and *strict* invariance) provide a proper goodness-of-fit level ($CFI > 0.90$, $TLI > 0.90$ and $RMSEA < 0.08$) and do not exceed the recommended cut off points for comparison of the increasingly restrictive models for $RMSEA$ ($\Delta < 0.015$), CFI ($\Delta < -0.01$) and TLI ($\Delta < 0.01$).

Validity of association with other variables

A regression analysis was performed to test to what extent the EECD-EF (independent variable) predicts the ISDS dimensions (dependent variable). The EECD-EF is proven to show a strong correlation with student *Satisfaction/enjoyment* subscale ($\beta = 0.29$, $p = 0.0001^{**}$), hence the more competent the teacher, the more enjoyment for students during PE class. The correlation is negative with *Boredom* subscale, but it cannot be proven that there is a prediction with this dimension because the values are not significant ($\beta = -0.07$, $p = 0.053$) (see Table 3).

Discussion

The purposes of this study were: to analyze the psychometric properties of the EECD-EF for secondary schools in the Mexican context; and to analyze the measurement factorial invariance according to *gender*. The psychometric properties of the EECD-EF instrument for the Spanish context were initially researched by Catano and Harvey (2011), who obtained an explained variance of 68%. The results obtained from the factorial structure in this study support reliability, validity and factorial invariance of the instrument adapted to the Mexican context. The CFA has been carried out with a single dimension (perception from students of the teachers' competency), similar to the hypothesized model in the Spanish context (Baena-Extremera et al., 2015) and the source model (Catano & Harvey, 2011).

In accordance with previous studies, it was found that the EECD-EF has adequate psychometric properties, as well as appropriate evidence on its internal structure according to the criteria established *a-priori*. The analyses of the present study yielded adequate internal consistency indices similar to those obtained in the analysis of the test in Spanish adolescents (Baena-Extremera et al., 2015). Likewise, the values obtained from the factor loadings are adequate, suggesting that the entirety of the EECD-EF items load on a single general factor. The unifactorial structure of the underlying model of the EECD-EF, proposed by Baena-Extremera et al. (2015) and supported

Table 2.
Goodness of fit indexes of the invariance models.

Model	χ^2	df	RMSEA [90% IC]	CFI	TLI	Model comparison	Δ RMSEA	Δ CFI	Δ TLI
M1	228.451	54	0.066 [0.057-0.074]	.955	0.939	-	-	-	-
M2	239.594	62	0.062 [0.054-0.070]	.954	0.946	2 versus 1	0.004	0.001	-0.007
M3	260.767	71	0.060 [0.052-0.067]	.951	0.950	3 versus 2	0.002	0.003	-0.004
M4	294.877	81	0.059 [0.052-0.067]	.944	0.950	4 versus 3	0.001	0.006	0.000

Note: χ^2 = Chi-square; df = degrees of freedom; RMSEA = root mean square error of approximation; CFI = confirmatory fit index; TLI = Tucker–Lewis index.

Table 3.
Simple linear regression of the EECD-EF (independent variable) predicts the subscales of the ISDS (dependent variable)

Subscales	F	β	p
Satisfaction/enjoy	66.71	0.29	0.0001**
Boredom	3.76	-0.07	0.053

Note: β =Standardized beta weights. **p < 0.01

by Baños, Baena-Extremera and Ortiz-Camacho (2019), was verified. The above means that it is possible to identify the perspective of Mexican high school students regarding their teachers’ teaching competencies in the subjects of PE, Mathematics, and Spanish.

Results obtained from the application of the MGCFA enabled the verification of evidence for factorial invariance of the EECD-EF scale in the *configurational, weak, strong* and *strict* models in terms of *gender* among Mexican secondary school students. Thus, it is confirmed that the underlying construct is equivalently understood by males and females. Likewise, comparisons of subjects based on mean scores are valid and unbiased (Dimitrov, 2010; Putnick & Bornstein, 2016). The results of the present research are similar to those obtained in the application of the scale in Spanish high school students (Martínez-Molina et al., 2020), so it is inferred that the scale retains its psychometric properties in different Hispanic cultural contexts. However, this would have to be assessed with a subsequent invariance study.

The ISDS was used for the external validity analysis because of the correlation between both variables and the potential practical applications from said correlation. The results show that EECD-EF positively predicts satisfaction with the PE class, however, there is not a significant correlation with boredom in PE. Other studies did find a negative prediction between EECD-EF and boredom in PE (Baena-Extremera et al., 2015). Also worth mentioning is how the PE teacher may be affected by the satisfaction with school in general, given that a student that considers the PE teacher to be competent will feel satisfied with school, otherwise, they will feel bored (Baños, Baena-Extremera, Ortiz-Camacho et al., 2019), a situation that may even unleash disruptive behaviors in the classroom (Baños, Baena-Extremera & Gra-

nero-Gallegos, 2019). The validation of this instrument for the Mexican context will make it easier to assess the PE teachers’ competencies within the Mexican education system. In this way, researchers will be able to analyze which competencies teachers must reinforce to increase chances of productive learning for the students. As underscored by Fortuin et al. (2015), it is essential that teachers acquire a large array of skills, not only in terms of knowledge, but also in emotional education, and class control and management.

It is worth highlighting the importance that the PETs’ competencies have within the curricula and development of adolescents, especially for the promotion of physical and healthy activities beyond the classroom (Stratton et al., 2008). Thus, understanding that PETs are able to consistently influence various aspects of students’ lives (Ruiz-Juan et al., 2017; Wallhead & Buckworth, 2004), this instrument may open new research paradigms for the PE field in the Mexican context. This is why it is important to underscore the contributions of this study for the improvement of PETs’ competencies in Mexico.

Conclusion

The present study demonstrates the adequate psychometric properties and evidence of construct validity of the EECD-EF for a sample of Mexican secondary school students. The results obtained support the unifactorial structure found in previous studies (i.e. Baena-Extremera et al., 2015), indicating that the instrument measures a particular phenomenon and not a variety of them; which facilitates the interpretations generated through the participants’ scores. Likewise, the unifactorial structure shows that it is invariant according to *gender* for the *configurational, weak, strong* and *strict* models, thus comparisons of mean scores between men and women are considered valid and unbiased (Dimitrov, 2010; Putnick & Bornstein, 2016). Consequently, the application of the EECD-EF is recommended for the measurement of identifying the students’ perspective in relation to teachers’ competencies because it’s a brief, inexpensive and easy-to-administer instrument.

Limitations and strengths

In terms of strengths, we underscore the sample's design, probabilistic and random by centers, stratified, multistage and by proportional affixation. In this way, results may be generalized for the state of Nuevo Leon, Mexico. In addition, the researched topic may greatly help generate solutions for the main issues faced by teachers every day. Also this instrument provides an easy and inexpensive way to measure students' perspective on the competencies of physical education, mathematics and Spanish teachers.

However, in spite of the sampling design and the interest for the research topic, some limitations must be considered. Even though the factorial invariance analysis was conducted according to sex, it would be interesting to broaden this research to other states of Mexico and different school levels to study the validity and reliability of the instrument with other subjects and be able to generalize results for the country regardless of the state and the student's academic level. Additionally, due to the fact that the results of the simple linear regression analysis contrast with studies applied to similar populations (Baena-Extremera et al., 2015), it is recommended that the instrument undergo a more robust regression analysis in order to understand the relations of the phenomenon in a more detailed manner.

Funding

Approval and funding for this study was obtained from the Secretary of Public Education and the Autonomous University of Baja California (ID number: 431/569/E).

References

- Ainley, J. & Carstens, R. (2018). *Teaching and Learning International Survey (TALIS) 2018 Conceptual Framework*. OECD Publishing. <https://doi.org/10.1787/799337c2-en>
- Aparicio, J.L., & Fraile, A. (2015). La evaluación de competencias interpersonales en la formación del profesorado de Educación Física a través de un programa de expresión corporal. *International Journal for 21st Century Education*, 2(2), 21-34. <https://doi.org/10.21071/ij21ce.v2i2.4266>
- Baena-Extremera, A., Granero-Gallegos, A., & Martínez-Molina, M. (2015). Validación española de la Escala de la Competencia Docente, Evaluación en Educación Física de secundaria. *Cuadernos de Psicología del Deporte*, 15(3), 113-122. <https://dx.doi.org/10.4321/S1578-84232015000300011>
- Baños, R., Baena-Extremera, A., & Granero-Gallegos, A. (2019). The Relationships between High School Subjects in terms of School Satisfaction and Academic Performance in Mexican Adolescents. *Int. J. Environ. Res Public Health*, 16(18), 3494-3507. <https://doi.org/10.3390/ijerph16183494>
- Baños, R., Baena-Extremera, A., & Ortiz-Camacho, M.D.M. (2019). Prediction of the satisfaction with the student life, based on teaching competence and satisfaction with the school. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.02506>
- Baños, R., Baena-Extremera, A., Ortiz-Camacho, M.M., Zamarripa, J., Beltrán A., & Juvera, J.L. (2019). Influencia de las competencias del profesorado de secundaria en los comportamientos disruptivos en el aula. *Espiral. Cuadernos del Profesorado*, 12(24), 3-10. <https://doi.org/10.25115/ecp.v12i24.2141>
- Baños, R., Barretos-Ruvalcaba, M., & Baena-Extremera, A. (2019). Protocolo de estudio de las variables académicas, psicológicas y de actividad física que influyen en el rendimiento académico de adolescentes mexicanos y españoles. *Espiral*, 12, 89-99. <https://doi.org/10.25115/ecp.v12i25.2480>
- Baños, R., Ortiz-Camacho, M.M., Baena-Extremera, A., & Tristán-Rodríguez, J.L. (2017). Satisfacción, motivación y rendimiento académico en estudiantes de Secundaria y Bachillerato. *Espiral. Cuadernos del Profesorado*, 10(20), 40-50. <https://doi.org/10.25115/ecp.v10i20.1011>
- Barba-Martín, R.A., Hernando-Garijo, A., Hortigüela-Alcalá, D., & González-Calvo, G. (2020). After nearly a decade of Bologna, have we really improved teaching quality? | *Espiral. Cuadernos del Profesorado*, 13(27), 97-108. <http://dx.doi.org/10.25115/ecp.v13i27.3186>
- Blázquez, D. (2013). *Diez competencias docentes para ser mayor profesor de Educación Física*. Barcelona: Inde.
- Bollen, K.A., & Long, J. (1994). *Testing structural equation models*. Newbury Park. Sage.
- Brown, T. (2015). *Confirmatory Factor Analysis for Applied Research*. Guilford Publications.
- Carretero-Dios, H., & Perez, C. (2007). Standards for the development and the review of instrumental studies: Considerations about test selection in psychological research. *International Journal of Clinical and Health Psychology*, 7(3), 863-882. http://www.aepc.es/ijchp/carretero_es.pdf
- Catano, V. M., & Harvey, S. (2011). Student perception of teaching effectiveness: development and validation of the Evaluation of Teaching Competencies Scale (ETCS). *Assess Eval High Edu*, 36(6), 701-717. <https://doi.org/10.1080/02602938.2010.484879>
- 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>
- Cheung, G.W., & Rensvold, R.B. (2002). Evaluating Goodness-Of-Fit Indexes for Testing Measurement Invariance. *Struct. Equation Model. A Multidisciplinary J.*, 9(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5
- Cochran, W, G. (1991). *Sampling Techniques*. John Wiley.
- Colman, A.M., Norris, C.E., & Preston, C.C. (1997). Comparing rating scales of different lengths: Equivalence of scores from 5-point and 7-point scales. *Psychological Reports*, 80(2), 355-362. <https://doi.org/10.2466/pr0.1997.80.2.355>
- de Winter, J.C.F., Gosling, S.D., & Potter, J. (2016). Comparing the pearson and spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological Methods*, 21(3), 273-290. <https://doi.org/10.1037/met0000079>
- Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, 43(2), 121-149. <https://doi.org/10.1177/0748175610373459>
- Fortuin, J., van Geel, M., & Vedder, P. (2015). Peer influences on internalizing and externalizing problems among adolescents: A longitudinal social network analysis. *Journal of Youth and Adolescence*, 44, 887-897. <https://doi.org/10.1007/s10964-014-0168-x>
- Gallego-Ortega, J.L., & Rodríguez-Fuentes, A. (2019) Percepciones del profesorado sobre competencias comunicativas de futuros maestros de educación. *Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte*, 19(75), 479-492. <https://doi.org/10.15366/rimcafd2018.71.005>
- Granero-Gallegos, A., Baños, R., Baena-Extremera, A., & Martínez-Molina, M. (2020). Analysis of misbehaviours and satisfaction with school in secondary education according to student gender and teaching competence. *Frontiers in Psychology*, 11(63). <https://doi.org/10.3389/fpsyg.2020.00063>

- Guillén-Gámez, F. D., & Perrino, M. (2020). Análisis Univariante de la Competencia Digital en Educación Física: un estudio empírico. *Retos: nuevas tendencias en educación física, deporte y recreación*, (37), 326-332. <https://doi.org/10.47197/retos.v37i37.72052>
- Hair, J.F., Black, W. C., Babin, B. J., & Anderson, R.E. (2019). *Multivariate Data Analysis*. Pearson New International Edition. Cengage Learning.
- Helsinki Declaration (2008). *World Medical Association*. <https://www.wma.net/es/politicas-post/declaracion-de-helsinki-de-la-amm-principios-eticos-para-las-investigaciones-medicas-en-seres-humanos/>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria. Conventional criteria versus new alternatives, *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1). <https://doi.org/10.1080/10705519909540118>
- INEGI (2019). *Características educativas de la población*. Instituto Nacional de Estadística y Geografía. <https://www.inegi.org.mx/temas/educacion/default.html#Tabulados>
- Jones-Jofré, J., Müller-Pérez, S., & Barraza-Gómez, F. (2022). Efecto de la asignatura de educación física en la percepción de los estudiantes frente al desarrollo de competencias genéricas. *Retos: nuevas tendencias en educación física, deporte y recreación*, 44, 496-503. <https://doi.org/10.47197/retos.v44i0.90654>
- Jöreskog, K.G., & Sörbom, D. (2003). *Structural equation modeling with the SIMPLIS command language*. Chicago: Scientific Software International.
- Liddell, T.M., & Kruschke, J.K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, 79, 328–348. <https://doi.org/10.1016/j.jesp.2018.08.009>
- Lidice, A., & Saglam, G. (2012). Using student evaluations to measure educational quality. *Procedia-Social and Behavioral Sciences*, 70, 1009-1015. <https://doi.org/10.1016/j.sbspro.2013.01.152>
- Lleixá, T., Torralba, M.A., & Ahrhao, S.R. (2010). Evaluación de competencias en Educación Física: Investigación-acción para el diseño de procedimientos de evaluación en la Etapa Primaria. *Movimiento*, 6(4), 33-51. <https://www.redalyc.org/pdf/1153/115316963003.pdf>
- Luna, E., Calderón, N., Caso, J. & Cordero, G. (2012). Desarrollo y validación de un cuestionario de evaluación de la competencia docente con base en la opinión de los estudiantes. En E. J. Cisneros-Cohernou, B. García-Cabrero, E. Luna y R. Marín (Coords.), *Evaluación de Competencias Docentes en la Educación Superior* (pp.119-158). Juan Pablos Editor.
- Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3), 519–530. <https://doi.org/10.1093/biomet/57.3.519>
- Marsh, H.W., Ginns, P., Morin, A.J.S., Nagengast, B., & Martin, A.J. (2011). Use of Student Ratings to Benchmark Universities: Multilevel Modeling of Responses to the Australian Course Experience Questionnaire (CEQ). *Journal of Educational Psychology*, 103(3), 733-748. <https://psycnet.apa.org/doi/10.1037/a0024221>
- Martínez-Mínguez, L., Moya, L., & Arnau-Sabatés, L. (2022). ECPP-FIM an assessment tool for evaluating the professional psychomotor skills within the initial teacher training: validity evidences. *Retos*, 43, 510-520. <https://recyt.fecyt.es/index.php/retos/index>
- Martínez-Molina, M., Granero-Gallegos, A., Baena-Extremera, A., & Baños, R. (2020). Invarianza factorial por sexo del Cuestionario para las Conductas Disruptivas y Escala de Evaluación de la Competencia Docente en Educación Física en estudiantes de secundaria. *Psychology, Society, & Education*, 12(2), 125-137. <https://doi.org/10.25115/psye.v12i2.3172>
- McClowry, S.G., Rodriguez, E.T., Tamis-LeMonda, C.S., Spellmann, M.E., Carlson, A., & Snow, D. L. (2013). Teacher/student interactions and classroom behavior: the role of student temperament and gender. *J. Res. Child Educ.* 27, 283–301. <https://doi.org/10.1080/02568543.2013.796330>
- Milfont, T.L., & Fisher, R. (2010). Testing measurement invariance across groups: Applications in cross. *International Journal of Psychological Research*, 3(1), 111–121. <https://doi.org/10.21500/20112084.857>
- Moreno-Murcia, J.A., Silveira, Y., & Beldando, N. (2015). Questionnaire evaluating teaching competencies in the university environment. Evaluation of teaching competencies in the university. *New Approaches In Educational Research*, 4(1), 54-61. <https://doi.org/0.7821/naer.2015.1.106>
- Navio, A. (2005). Propuestas conceptuales en torno a la competencia profesional. *Revista de Educación*, 337, 213-234. <https://redined.meecd.gob.es/xmlui/bitstream/handle/11162/68040/00820053000213.pdf?sequence=1>
- OECD (2013). *Teaching and Learning International Survey TALIS 2013. Conceptual Framework*. https://www.oecd.org/education/school/TALIS%20Conceptual%20Framework_FINAL.pdf
- Ojeda-Nahuelcura, R., Carter-Thuillier, B., López-Pastor, V., Fuentes-Nieto, T., & Gallardo-Fuentes, F. (2022). Evaluación de competencias genéricas en profesores de Educación Física. *Retos: nuevas tendencias en educación física, deporte y recreación*, (43), 521-532. <https://doi.org/10.47197/retos.v43i0.88796>
- OXFAM NOVID (2011). *Educadores de calidad: Estudio internacional sobre las competencias y estándares para docentes*. Internacional de la Educación. <https://download.ei-ie.org/Docs/WebDepot/Quality%20Ed%20Study%20ES.PDF>
- Palacios, A., López-Pastor, V., & Fraile, A. (2019). Cuestionario de percepción de competencias docentes de Educación Física. *Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte*, 19(75), 445-461. <https://doi.org/10.15366/rimcafd2019.75.005>
- Perea R.L., & Abello, C. M. (2022). Competencias digitales en estudiantes y docentes universitarios del área de la educación física y el deporte. *Retos: nuevas tendencias en educación física, deporte y recreación*, (43), 1065-1072. <https://dialnet.unirioja.es/servlet/articulo?codigo=8106411>
- Rivera, A., & Alfageme-González, M.B. (2019). Teacher Training for In-Service Secondary School Teachers in Mexico. *Revista Educación*, 43(1), 30-43. <https://doi.org/10.15517/revedu.v43i1.28446>
- Ruiz-Juan, F., Baena-Extremera, A., & Baños, R. (2017). Nivel de actividad deportiva en el tiempo libre desde las etapas de cambio y motivación en estudiantes de Costa Rica, México y España. *Cuadernos de Psicología del Deporte*, 17(2), 53-64. <https://revistas.um.es/cpd/article/view/301891>
- Salcines-Talledo, I., González-Fernández, N., Ramírez-García, A., & MartínezMínguez, L. (2018). Validación de la Escala de Autopercepción de Competencias Transversales y Profesionales de Estudiantes de Educación Superior. Profesorado. *Revista de Currículum y formación del Profesorado*, 22(3), 31-51. <https://doi.org/10.30827/profesorado.v22i3.7989>
- Secretaría de Educación Pública (2017). *Perfiles, parámetros e indicadores para docentes y técnicos docentes en Educación Básica*. SEP. http://www.sev.gob.mx/servicio-profesional-docente/files/2017/01/PPI_INGRESO_EB_2017_2018.pdf
- Stratton, G., Fairclough, S.J., & Ridgers, N.D. (2008). Physical activity levels during the school day. En A.L. Smith y S.J.H. Biddle (Eds), *Youth Physical Activity and Sedentary Behavior. Challenges and Solutions* (pp. 321–350). Human Kinetics.
- Putnick, D.L., & Bornstein, M.H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71-90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Tejada, J. (2005). El trabajo por competencias en el prácticum: cómo organizarlo y cómo evaluarlo. *Revista Electrónica de Investigación Educativa*, 7(2). <http://redie.uabc.mx/vol7no2/contents-tejada.html>
- Tul, M., Leskosek, B., Jurak, G., & Kovac, M. (2015). Perceived importance of slovenian physical education teachers' professional competencies. *HU Journal of Education*, 30(1), 268-281. <http://efdergi.hacettepe.edu.tr/yonetim/icerik/makaleler/35-published.pdf>
- Vandenberg, R.J., & Lance, C.E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Res. Methods*, 3(1), 4–70. <https://psycnet.apa.org/doi/10.1177/109442810031002>
- Wallhead, T., & Buckworth, J. (2004). The role of physical education in the promotion of youth physical activity. *Quest*, 56, 285-301. https://www.researchgate.net/publication/234762436_The_Role_of_Physical_Education_in_the_Promotion_of_Youth_Physical_Activity