

General Musical Self-Efficacy Scale: Adaptation and validation of a version in Spanish

Journal:	<i>Psychology of Music</i>
Manuscript ID	POM-21-2113.R2
Manuscript Type:	Original Empirical Investigations
Keyword:	Musical self-efficacy, Self-efficacy for learning, Self-efficacy for performing, Music students, Scale adaptation, Validation
Abstract:	<p>This article presents the process of translation into Spanish, adaptation, and validation of the General Musical Self-Efficacy Scale by Ritchie and Williamon (2007, 2011a). After having carried out reverse translation and a preliminary study, the scale was tested on a sample of 668 students enrolled in six Spanish music academies (secondary-school and university level). Our results corroborated the two original subscales: one for self-efficacy of learning ($\alpha = .792$ and $\alpha = .734$), and the other for self-efficacy for public performance ($\alpha = .773$ and $\alpha = .780$), after having removed one item in each subscale on the basis of internal consistency indicators and factorial analysis. The subscales we thereby obtained likewise presented good temporal stability ($r = .513$ and $.539$); they both correlated with performance anxiety (stage fright); both correlated more highly with the public performance subscale. Validation data signaled higher scores obtained by students enrolled in university-level music academies, by students who had started music training at an earlier age, for the boys at secondary level, and for the younger group of students at secondary level. To summarize, this version of the General Musical Self-Efficacy Scale is a valid, reliable tool for measuring self-efficacy in Spanish music students.</p>

General Musical Self-Efficacy Scale: Adaptation and validation of a version in Spanish

Abstract

This article presents the process of translation into Spanish, adaptation, and validation of the *General Musical Self-Efficacy Scale* by Ritchie and Williamon (2007, 2011a). After having carried out reverse translation and a preliminary study, the scale was tested on a sample of 668 students enrolled in six Spanish music academies (secondary-school and university level). Our results corroborated the two original subscales: one for self-efficacy of learning ($\alpha = .792$ and $\alpha = .734$), and the other for self-efficacy for public performance ($\alpha = .773$ and $\alpha = .780$), after having removed one item in each subscale on the basis of internal consistency indicators and factorial analysis. The subscales we thereby obtained likewise presented good temporal stability ($r = .513$ and $.539$); they both correlated with performance anxiety (stage fright); both correlated more highly with the public performance subscale. Validation data signaled higher scores obtained by students enrolled in university-level music academies, by students who had started music training at an earlier age, for the boys at secondary level, and for the younger group of students at secondary level. To summarize, this version of the General Musical Self-Efficacy Scale is a valid, reliable tool for measuring self-efficacy in Spanish music students.

Keywords

Musical self-efficacy, Self-efficacy for learning, Self-efficacy for performance, Music students, Scale adaptation, Validation.

Many recent studies in the area of education have pursued the main goal of determining which factors have an influence on academic success (Zimmerman, 2000). In music education, the Self-Regulated Learning theory has become one of the most useful models for the explanation of specific successes achieved by students on different academic levels (McPherson & McCormick, 2000; Varela et al., 2016). In it, self-efficacy (Bandura, 1997) emerges as one of the psychological constructs with the greatest relevance, but a greater amount of research is still required to empirically support its development (Hendricks, 2014). One of the motives that limit its development is precisely the lack of appropriate evaluation tools. To date, no valid tools had been developed for Spanish-speaking populations. We thus attempted to adapt into Spanish one of the most recent tools developed in the English language: the *General Musical Self-Efficacy Scale* by Ritchie and Williamon (2007, 2011a).

A series of studies in the area of music have demonstrated the relationships between beliefs of self-efficacy, musical achievement, and the strategies and processes involved in the self-regulated musical learning model. Several classic studies led by Gary McPherson were published at the onset of the 21st century (McCormick & McPherson 2003; McPherson & McCormick, 2000, 2006), in which the authors observed the capacity of self-efficacy to explain or predict musical achievement in public performance, while still regarding it as part of a motivational model. A systematic review of this subject can be found in Varela et al. (2016). More recent studies have attempted to offer further support for the relationships between musical achievement and self-efficacy, as well as with the components of the self-regulated learning model (Hendricks, 2014, Hewitt, 2015, Miksza and Tan, 2015, Ritchie and Williamon, 2013).

To approach the problem of evaluating self-efficacy, it is necessary for us to pay heed to a further aspect of this construct: self-efficacy can be restricted to specific areas and domains (Hendricks, 2014; Ritchie & Williamon, 2011a). Although a general dimension of self-efficacy has been used for purposes of analyzing its role in the performance anxiety experienced by musicians (Orejudo et al., 2017), Bandura (2006) himself established that self-efficacy is truly predictive when it is applied to restricted fields with a sufficient degree of concretion that allows us to identify specific behaviors. In the area of music, this requires that we delimit and define the concrete areas that require musical competency. A series of alternatives have been elicited, of which certain ones have one sole musical self-efficacy factor, and others have several. As a consequence, musical self-efficacy scales with one sole factor have been designed: for example, in the pioneer studies by McCormick and McPherson (2003), and McPherson and McCormick (2006), or, more recently, by Upitis et al. (2017) in their scale entitled *Self-efficacy: Personal competency beliefs*, associated with self-efficacy, which include public performance elements, musical practice elements, and the social recognition of the competency of musicianship.

Almost all of the scales which have been developed to evaluate self-efficacy include certain elements associated with musical performance, but can likewise propose a series of other components. Thus, for instance, Ritchie and Williamon (2007, 2011a) establish a distinction between self-efficacy for learning and self-efficacy for performing. According to Ritchie and Williamon (2011a) “learning focuses predominantly on the *acquisition* of skills and knowledge, with various self-regulated learning strategies underpinning this, whereas performance typically centres on *implementation*” (p. 330). However, it is also possible that factors involved in one or the

other type of self-efficacy can be quite distinct from one another. Ritchie and Williamon (2013) highlighted the greater degree of influence exerted by metacognitive and self-regulating processes in self-efficacy for learning; these, however, are barely relevant in the musical performance situation. On the other hand, performance anxiety (stage fright) is a key construct which plays a role in levels of self-efficacy for performing, but is much less involved in self-efficacy for learning (Zarza et al., 2020). Papageorgi et al. (2013) likewise established the same distinction between self-efficacy for learning and for performing, noting the latter's close relation with stage fright (performance anxiety). Various other authors have underscored further aspects of musical competency in their respective scales, for example Randles (2011), Uptis et al. (2017) or Zelenak (2010) delve deeper into the components of self-efficacy or Ekinici (2014), Kurtuldu and Bulut (2017), Orton and Pitts (2019) or Watson (2010) develop specific scales related to musical or instrumental styles.

Another aspect also included in self-efficacy scales has to do with the answer format and the number of items they should contain, a criterion regarded as key (Hendricks, 2014). The first studies by McPherson and McCormick (2003) and McCormick and McPherson (2006) featured few items, whereas Ritchie and Williamon (2007, 2011a) feature a scale format with a wide variety of items. In terms of answer format, most authors have opted for a Likert scale, although Bandura proposes the use of an uninterrupted 100-point scale with 10-unit intervals (Bandura, 2006).

This all implies the need to gain further knowledge about musical self-efficacy; this study's main objective is to adapt and validate a Spanish-language version of a tool that shall permit us to analyze levels of self-efficacy perceived by music students in institutions or musical learning in Spain. After having exhaustively analyzed the tools in existence, we chose to translate, adapt, and validate the *General Musical Self-Efficacy Scale* conceived by Ritchie and Williamon (2007, 2011a), which is coherent with Bandura's Social Cognitive Theory (Bandura, 1977). Ritchie and Williamon's scale is based on the *Specific General Self-Efficacy Scale* (Sherer et al., 1982). Ritchie and Williamon (2007) redacted the 17 items of that scale and adapted them to the context of music education. In view of the construct's specificity, they developed two subscales (musical self-efficacy for learning and musical self-efficacy for performing), with a 7-point Likert format (from 1: "disagree" to 7: "agree") and 11 items per subscale. The scale was validated in 2011 on a sample of 250 music conservatory and music university students: one of its subscales the learning scale, was applied to primary education students (Ritchie & Williamon, 2011, 2013), an option similar to that which was carried out by Urruzola and Bernaras (2020), who applied it to Spanish students of the same age: 8 to 12 years old.

The adaptation we carried out in this study was accomplished by testing the scale via a procedure of reverse translation, after which the scale was applied to a sample of conservatory students in Spain (secondary and university levels), over a much wider age range: from teenagers to adults. To obtain proof of convergent and divergent validity, we analyzed the self-efficacy relations with the Kenny Music Performance Anxiety Inventory (K-MPAI) by Kenny et al. (2004) in its Spanish version (Zarza et al., 2016). Other evidences of criterion validity were obtained by comparing academic level, gender and age. We applied CFA to analyze construct validity, with factorial invariance in the two educational levels under study (secondary-level and university-level). The guiding hypothesis for this study was that we would be able to obtain a scale of musical self-efficacy with a structure similar to the original: i.e., with two factors; furthermore, that the obtained scale would have concrete correlations with K-MPAI

(Performance Anxiety Inventory), regarding which students would vary in function of their academic level. Also, as in previous cases, we expected to find a certain degree of variance according to gender.

Method

Participants

The sample was made up of 668 students enrolled in six institutions of musical learning in different cities; 47.1% of them were male, 52.9% female. In terms of academic level, 70.8% were enrolled in advanced secondary-school level music schools (*conservatorios profesionales*), and 29.2% were enrolled in university-level music academies (*conservatorios superiores*). The percentage of girls and boys is not statistically significantly different among the two academic levels ($\chi^2 = .936, p = .626$). All in all, the sample comprises students of ages ranging from 11 to 61 years old, with a median age of 18.13 ($S.D. = 5.47$), and an important difference between those enrolled in *conservatorios profesionales* [$\bar{x} = 16.82$ ($S.D. = 5.31$)] and *conservatorios superiores* [$\bar{x} = 21.43$ ($S.D. = 4.57$)]. The range of musical specialties is broad: bowed strings (25.6%), woodwinds (25.1%), keyboards (16.9%), brass (14.8%), plucked strings (8.7%), percussion (4.3%) and other specialties (4.3%). The majority are either predominantly orchestral instruments (67.8%) or predominantly solo instruments (26.5%), but other types appear as well, such as orchestral conducting, traditional folklore instruments, jazz, and early music (5.5%).

Variables and tools

Data were gathered via an ad hoc battery of questionnaires including questions regarding sociodemographic and educational variables (age, gender, academic year, and age of musical onset, musical instrument), the Spanish-language adaptation of the *General Musical Self-Efficacy Scale* by Ritchie and Williamon (2007, 2011a), and the subscale of cognitions in the Spanish-language version of K-MPAI, the *Kenny Music Performance Anxiety Inventory* (Zarza et al., 2016).

The *General Musical Self-Efficacy Scale* by Ritchie and Williamon (2007, 2011a) is a 7-point Likert-type questionnaire (1: “disagree” – 7: “agree”), made up of 22 items grouped into two subscales: self-efficacy for learning and self-efficacy for performing. In each one of them, six items were reverse coded: Items Nos. 2, 4, 5, 8, 10, and 11 under the learning factor, and items 2, 3, 4, 6, 7, and 8 under the performing factor. The Spanish version can be found in supplementary material.

In order to evaluate performance anxiety (stage fright), we used the Cognitions subscale of the Spanish-language adaptation of KMPAI, the *Kenny Music Performance Anxiety Inventory* (Zarza et al., 2016). The 9 items of this scale measure the specific factor of anxiety associated with proximal performance concerns ($\alpha = .854$).

Procedure

After having received an affirmative response from the above-cited institutions of musical learning, we proceeded to gather the data in person, on the premises. The research team visited the academies in order to operate *in situ*, with the exception of the Conservatories of Pamplona and Elche, where a local professor was entrusted with the

task of administering and gathering the questionnaires. Students participated on a voluntary, anonymous basis. Consent was obtained from participants, in the case of children under fourteen from families. Ethics clearance was granted by Research Ethics Committee of the Autonomous Community of Aragon.

The procedure of translation and adaptation of the scale, as well as our statistical analysis, is described in the Supplementary Material section, which also includes the scale in its Spanish version.

Results

Descriptive analysis and reliability

An initial analysis of the item scores yields medium-high values for the subscale “self-efficacy for learning” ($\bar{x} = 4.92\text{--}6.17$), which are slightly higher than those yielded by the subscale “self-efficacy for performing” ($\bar{x} = 4.46\text{--}5.79$). None of the items on any of the two scales has values that are distant from the rest in its subscale. Items with lower scores have normal distribution according to values for skewness and kurtosis, whereas four items in the subscales “self-efficacy for learning” (5, 8, 10, and 11) and two items of the subscale “self-efficacy for performing” (6 and 11) present with slight negative skewness, with values below -1 .

Table 1 displays the internal consistency results of the two self-efficacy subscales according to educational level. The sample of non-university students has higher internal consistency values. In both cases, the elimination of an item (Item No. 9 in learning and Item No. 10 in performing) improves the subscales’ reliability. The elimination is also justified by the item-total correlations, since these are the two items that have the least correlation with the others: $r = .108$ (Item No. 9 in learning) and $r = .064$ (Item No. 10 in performing). The remaining correlations are greater than $.30$.

Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA)

After descriptive analysis and reliability testing, the next task was to establish the number of factors that should be included in the exploratory model ($n = 336$). Results varied slightly, according to whether we included all 22 original items of the scale, or only the 20 items retained after reliability analysis. With 22 items, parallel analysis determined three factors (table 2), but with 20 items, two are sufficient. Thus EFA confirms the scale’s bidimensionality. It is nevertheless important to point out that the two factors only explain 34.47% of the variance of the 20 items.

With CFA we explored several analysis options. On the one hand, we validated the factorial structure that emerged from EFA with a two-factor model: one structure which grouped items related to self-efficacy for learning, and the other structure grouping items related to self-efficacy for performance. All regression coefficients were significantly different from zero and present values lying between $.30$ and $.60$, thereby confirming the scale’s communalities (table 2). We also confirmed that it is optimal to remove the two items which had been less valid and reliable in the previous analyses (Item 9 in self-efficacy for learning, and Item 10 in self-efficacy for performance). Both items had very low regression weights, lying below $.30$. As shown in Supplementary material (table 4), the confirmatory model’s goodness of fit improved significantly in Model 2, containing 20 items.

On the other hand, we tested two analysis strategies based on the explanatory factorial structure which the scale's original authors had already noted. It consists in an independent analysis of each scale with two-factor models in each one: one for direct items and the other for inverse items. Thanks to this strategy, we found an optimal structure for the "self-efficacy for learning" scale, with a two-factor model for direct items on the one hand, and for inverse items on the other (table 4, Model 3, in Supplementary material). In this model, the two factors would correlate ($r = .544$). A unifactorial model of this scale, considered independently, adjusts less well (Model 4). Analysis of regression weights (table 2) shows that in Factor 2, which groups the inverse items, certain of them have values below .50, namely Items 2 and 4. In Factor 1, all weights are over .50, but not all of them are over .60.

Regarding the "self-efficacy for performance" subscale, the two-factor model (Model 5) yields a less satisfactory result than the one it yields for the "self-efficacy for learning" subscale; it is better, at any rate, than the model for that sole subscale with just one factor (Model 6). The two factors for Model 5 are highly correlated with one another ($r = .635$). Analysis of regression weights (table 2) shows that three items pertaining to Factor 2 (Perf_2, Perf_3 and Perf_4) are below .50.

In a final step, we tested a model for the entire scale, combining the two subscales in one sole analysis, but with four factors: two for each scale. This model (Model 7 in Supplementary Material, table 4) has moderate goodness of fit values, as in the two previous cases, but the most relevant aspect is the correlation pattern among the four factors. Those that separately group the items of each individual scale are correlated amongst themselves (F1 Learn and F2 Learn = .544; F1 Perform and F2 Perform = .623), but those that group the items according to response type, with or without inverse scores, correlate higher than the rest (F1 Learn and F1 Perform = .805; F2 Learn and F2 Perform = .843).

Validation. Relationship with the other variables

A final indication of validity can be found in the relationships with the other variables in this study. Thus, as a criterion of validity, we display the differences obtained in function of four variables: the type of institution and academic year ("Profesional", pre-university: 6 academic years, vs. "Superior", university-level: 4 academic years), gender (male and female), current age (in four groups: ages 11 to 15, 16 to 17, 18 to 20, and over 20 years old)- The only differences that appear on the "self-efficacy for learning" subscale are in function of age (table 2): 16-to-17-year-old students score significantly lower than the other groups (self-efficacy for learning = 54.73). All variables present significant differences on the "self-efficacy for performing" subscale. University-level students score higher than non-university-level students (52.99 vs. 50.85). In terms of academic year, the results reiterate the differences between types of learning institution: within each one of them, they reflect progress from one academic year to the next. In terms of academic year, there are no differences on the "self-efficacy for learning" subscale ($F_{9,608} = 1.429$, $p = .172$, $\eta^2 = .021$), but significant differences do appear on the "self-efficacy for performance" subscale ($F_{9,613} = 2.575$, $p = .006$, $\eta^2 = .036$). Figure 1 displays the mean scores of each academic year. University-level students score notably higher on "self-efficacy for performing" from their second academic year on; the lowest scores are present in the two last years of *conservatorio profesional* (pre-university) and the first year of *conservatorio superior* (university level). Scores are higher in males vs. females (53.01

vs. 50.23); scores are lower in 16-to-17-year-old students compared to the rest (49.50). Nevertheless, size effects are not particularly large: the most outstanding ones are gender-related.

We subsequently present the final indicator of validity: the relationships between the two self-efficacy subscales and the cognitions scale from the K-MPAI performance anxiety inventory. In both cases, the correlation is statistically significant, and negative (-.448 with the learning subscale and -.685 with the performing subscale). In both cases the correlation is inverse: in other words, the more self-efficacy the student perceives, the less anxiety they feel about performing. Correlations between the two self-efficacy subscales are likewise statistically significant and considerably high ($r = .645$): the two subscales might thus have a variability element in common. To test this hypothesis, we carried out partial correlations among the three variables. The correlation between K-MPAI cognitions and self-efficacy for learning ($r = -.043$, $p = .360$) is no longer statistically significant when the variance shared with self-efficacy for performance is eliminated, whereas the correlation between K-MPAI cognitions and self-efficacy for performing is maintained at a similar level when their shared variance with self-efficacy for learning is eliminated ($r = -.574$, $p < .001$). The correlation between the two types of self-efficacy is maintained when their shared variance with K-MPAI performance anxiety cognitions is eliminated, resulting in a value slightly below direct correlation ($r = .492$, $p < .001$ vs. $r = .645$).

Reliability over time

The two factors resulting from the adaptation process obtain acceptable indices, but with a correlation among the two points in time of $r = .513$ for the learning subscale and of $r = .539$ for the performance subscale. As indicated in Supplementary Material, we obtained these data on a subsample of 75 participants.

Discussion and conclusions

This study’s main objective was to translate, adapt, and validate a Spanish-language version of Ritchie and Williamon’s *General Musical Self-Efficacy Scale* (Ritchie and Williamon 2007, Ritchie and Williamon 2011a). The result is a valid, reliable tool that can be applied to students in Spain. To achieve this, we followed expert recommendations for the adaptation and validation of scales (Abad et al., 2011; Byrne, 2012; Muñiz et al, 2013): to these purposes we tested the scale on a wide, representative sample of Spanish music students. We included the two most important types of institutions of musical learning which apply the official curriculum: 1) *conservatorios superiores*, university-level music academies for older students who want to embark on a professional career in music; 2) *conservatorios profesionales*, which impart training designed to prepare students for access to the former, and in which other students are likewise enrolled, but who practice music for reasons of leisure or general education and formation.

Our adaptation process had partially satisfactory results, although the original scale we designed did present certain difficulties. The 11 original items on each of the two subscales in the Spanish version had to be reduced to 10, eliminating one item on each subscale. The eliminated items shared similar textual content: “the thought of failing in this performance leads me to work harder on preparing it” (Item No. 9 in Learning), and “the thought of failing in this performance makes me work harder” (Item

No. 10 in Performing). A first explanation of this result could lie in their translation: perhaps the catastrophic connotation of the word *fracasar* (for “failing”) could be the reason why these items do not function like the rest. But the problem could also lie in the original scale. When Ritchie and Williamon validated their scale on conservatory students (2011a), this item presented the same difficulties in the “self-efficacy for performing” subscale. They also eliminated a further item. In our case, the elimination of two scale items is justified by the internal consistency analysis, as well as by EFA and CFA analysis. Our age sample is broader and includes a wider swath of population; thus, our version does better to leave these items out. Their content might be misunderstood, since it implies that we are anticipating a necessary failure that obliges the student to invest greater effort.

After having eliminated those two items, the two subscales have internal consistency values over .70, which, without being as optimal as those of other scales, can nevertheless be regarded as reliable. Those values are slightly lower than those found by other authors in conservatory students: $\alpha = .82$ for self-efficacy in learning and $\alpha = .78$ for self-efficacy in performance (Ritchie & Williamon, 2011a). The values we obtained improve when educational level is taken into account: concretely, self-efficacy for learning works better with non-university-level students ($\alpha = .792$), whereas self-efficacy for performing works better with university-level students ($\alpha = .780$). The scale likewise has good temporal stability, which adds to its overall reliability. On younger students we only applied the self-efficacy subscale for learning, which has proven to be valid and reliable for those ages. Thus, Ritchie and Williamon (2013) found good internal consistency ($\alpha = .87$) for the English-language version with 11 items, and Urruzola and Bernaras (2020) obtained good results with Spanish students ($\alpha = .70$); however, the latter did not provide more detailed information about the items.

CFA analysis also brings out certain characteristics associated with the scale’s reverse-coded items. For the two subscales, we found that a bifactorial model associated with the reverse-coded items presents good fit values and improves the assumption of a single-factor scale for each factor: this, in turn, implies that the answer format plays an important role in the scale’s overall validity. This situation particularly affects the subscale “self-efficacy for performing”, in which the factorial weights of several items are over .40 but do not reach the optimal value of .60. This, in turn, leads to increased problems of fit in the models, as would be shown in certain indicators such as *CFI* and *TLI*, which would then lie below the optimal values, although further fit values of the models such as the χ^2/df ratio, *RMSEA*, and *SMSR* would still obtain adequate values. Factorial invariance analysis likewise reveals certain differences in function of the characteristics of the learning institution, from which we derive the possibility of making adjustments in its use, in function of the type of students under analysis.

Regarding the validation study, the tool has demonstrated its capacity to discriminate among student profiles, particularly regarding self-efficacy for performance, which is higher among university-level students who are oriented toward making public musical performance their profession. Despite only having obtained significant mean average differences in the performing factor, university-level students (*conservatorios superiores*) had higher scores in both self-efficacy factors (learning and performing) compared with their non-university-level colleagues (*conservatorios profesionales*). These results could be due to a curriculum in which the relationship between personal study time and class time has considerably increased, as well as the progressive increase in activities related with performing in public, which, in turn can

lead students to have improved self-efficacy perceptions thanks to their own mastery experiences (in terms of quantity), their vicarious experiences, and verbal persuasion. At any rate, these differences do not have a very large effect size, which could be due to the fact students do not tend to compare themselves with students studying at other levels, but with their colleagues enrolled in the same institution of musical learning. This could thus entail that the differences are not very large.

The subscale “self-efficacy for learning” displays no significant differences between university-level and non-university-level, but the means profile in Figure 1 displays several peaks similar to those observed in the subscale “self-efficacy for performance”. It is interesting to observe that students in the last years of non-university *conservatorio profesional*, approx. 16 to 17 years old, are those who have the lowest self-efficacy profile. This is in line with previous studies (Papageorgi et al., 2007; Orejudo et al., 2020; Zubeldia et al., 2018), in which students of the same age perceived a lesser degree of social support in music-oriented activities, and also displayed a greater amount of performance anxiety. The simplest way to interpret these findings is to note that students at that age are required to decide about their professional future, and they have a considerable study load to face in non-musical subjects as well, since they are finishing high school in the two last *bachillerato* courses and will have to pass the *selectividad* exam which grants them access to university; many of them will no longer opt to continue studying music in a *conservatorio superior*. It could be, in this situation, that the evaluation of one’s self-efficacy is conditioned by those professional choices. At the same time, the subscale “self-efficacy for performance” yields higher scores in males, as well as in those students who started learning music before they were seven years old.

Significantly lower scores obtained by females in the levels of self-efficacy for musical performance confirm findings from previous studies (Hendricks, 2009, 2014; Hendricks et al., 2016; Nielsen, 2004; Wehr-Flowers, 2006) as well as from research in related psychological constructs, in which women have higher levels of performance anxiety (Orejudo et al., 2017; Papageorgi et al., 2007). The fact that no significant mean average differences between men and women were found in the factor “self-efficacy for learning” points to the influence of other variables on the performance factor that will need to be analyzed in future studies. A further finding from the validation procedure is thereby worth mentioning: the higher relationships with performance anxiety are found in self-efficacy for performance. Papageorgi et al. (2010) obtained a similar result, but with a lower correlation ($-.33$), perhaps due to the fact of having used other, less specific tools. As previously mentioned, achieving a high degree of competency in this domain requires a good handling of the anxiety associated with performance situations. This value is higher than in the case of self-efficacy for learning, as is indeed to be expected: although it would normally not have to be affected by performances, it does indeed share a variance component with performance situations, since, as commented above, the achievement of a high degree of competency in performance situations requires a good handling of the anxiety associated with them. Notwithstanding, partial correlation analysis yields the expected result: a lesser degree of influence of performance anxiety on self-efficacy for learning. This thereby provides new evidence for the conceptual distinction between the two types of musical self-efficacy. Ritchie and Williamon (2011a) had already shown this, although they carried out independent analyses of the two subscales. Our study thus provides new proof of construct validity: not only in relation with performance anxiety, but also as construct validity in function of academic year and age.

To summarize, the scale elaborated and tested herein is a valid, reliable tool for use in Spanish-language contexts, particularly as long as no other alternatives are found to resolve certain difficulties it still presents. Our study has certain limitations. The main one would be associated with sample adequacy. Our sample only featured students who had already achieved a certain degree of training: only students enrolled in *conservatorios profesionales* and *conservatorios superiores*, but not in elementary education, amateur musicians, or students receiving training in non-government-approved private music schools. Ritchie and Williamon (2011b), as well as Urruzola and Bernaras (2020), have indeed shown the learning subscale's validity for earlier ages than those featured in our study. In this study we maintained the original scale's response system: 7 discrete points. One could test whether results are maintained or perhaps even improved by applying other score systems with scales of up to 100 intermediate points, as suggested by Bandura (2006, 2012) and by other authors (Hendricks, 2014). Another aspect not taken into consideration in our validation study is the possible influence of contextual variables associated with the specific musical instrument or with the learning institution as elements playing a key role in self-efficacy, as noted, for instance, by Papageorgi et al. (2010).

References

- Abad, F. J., Olea, J., Ponsoda, V., & García, C. (2011). *Medición en Ciencias Sociales y de la Salud*. Síntesis.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191-215. doi: 10.1037/0033-295X.84.2.191
- Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares, & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307-337). Information Age Publishing.
- Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *Journal of Management*, 38(1), 9-44. doi: 10.1177/0149206311410606
- Byrne, B. M. (2012). *Structural Equation Modeling with Mplus. Basic Concepts, Applications, and Programming*. Routledge.
- Ekinci, H. (2014). Validity and reliability study of the self-efficacy scale in rendering piano education to children of 6-12 years. *Educational Research and Reviews*, 9(18), 691-698. doi: 10.5897/ERR2014.1788
- Hendricks, K. S. (2009). *Relationships between the sources of self-efficacy and changes in competence perceptions of music students during an all-state orchestra event*. Unpublished doctoral thesis. University of Illinois. <https://www.ideals.illinois.edu/handle/2142/85814>
- Hendricks, K. S. (2014). Changes in Self-efficacy beliefs over time: Contextual influences of gender, rank-based placement, and social support in a competitive orchestra environment. *Psychology of Music*, 42(3), 347-365. doi: 10.1177/0305735612471238

- Hendricks, K. S., Smith, T. D., & Legutki, A. R. (2016). Competitive comparison in music: influences upon self-efficacy beliefs by gender. *Gender and Education*, 28(7), 918-934. doi: 10.1080/09540253.2015.1107032
- Hewitt, M. P. (2015). Self-efficacy, self-evaluation, and music performance of secondary-level band students. *Journal of Research in Music Education*, 63(3), 298-313. doi: 10.1177/0022429415595611
- Kenny, D. T., Davis, P., & Oates, J. (2004). Music performance anxiety and occupational stress amongst opera chorus artists and their relationship with state and trait anxiety and perfectionism. *Journal of Anxiety Disorders*, 18(6), 757-777. doi: 10.1016/j.janxdis.2003.09.004
- Kurtuldu, M. K., & Bulut, D. (2017). Development of a self-efficacy scale toward piano lessons. *Educational Sciences: Theory and Practice*, 17(3), 835-857.
- McCormick, J., & McPherson, G. E. (2003). The role of Self-Efficacy in a Musical Performance Examination: An exploratory structural equation analysis. *Psychology of Music*, 31(1), 37-51. doi: 10.1177/0305735603031001322
- McPherson, G. E., & McCormick, J. (2000). The contribution of motivational factors to instrumental performance in a music examination. *Research Studies in Music Education*, 15(1), 31-39. doi: 10.1177/1321103X0001500105
- McPherson, G. E., & McCormick, J. (2006). Self-efficacy and music performance. *Psychology of Music*, 34(3), 322-336. doi: 10.1177/0305735606064841
- Miksza, P., & Tan, L. (2015). Predicting collegiate wind players' practice efficiency, flow, and self-efficacy for self-regulation: An exploratory study of relationships between teachers' instruction and students' practicing. *Journal of Research in Music Education*, 63(2), 162-179. doi: 10.1177/0022429415583474
- Muñiz, J., Elosua, P., & Hambleton, R. K. (2013). Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema*, 25, 151-157. doi: 10.7334/psicothema2013.24
- Nielsen, S. G. (2004). Strategies and self-efficacy beliefs in instrumental and vocal individual practice: A study of students in higher music education. *Psychology of Music*, 32(4), 418-431. doi: 10.1177/0305735604046099
- Orejudo, S., Candela, C., Casanova, O., & Cuartero, L. M. (2020). A social support scale for music students in music schools, academies, and conservatories: An adaptation into Spanish and a factorial invariance study. *Psychology of Music*. doi: 10.1177/0305735620968626
- Orejudo, S., Zarza, F. J., Casanova, O., Rodríguez, C., & Mazas, B. (2017). The relation of music performance anxiety (MPA) to optimism, self-efficacy and sensitivity to reward and punishment: Testing the theory of personal vulnerability on a sample of Spanish music students. *Psychology of Music*, 45(4) 570-583. doi: 10.1177/0305735616674791

- Orton, A., & Pitts, S. E. (2019). Adolescent perceptions of singing: Exploring gendered differences in musical confidence, identity and ambition. *Music Education Research, 21*(1), 40-51. doi: 10.1080/14613808.2018.1507019
- Papageorgi, I., Hallam, S., & Welch, G. (2007). A conceptual framework for understanding musical performance anxiety. *Research Studies in Music Education, 28*, 83-107. doi: 10.1177/1321103X070280010207
- Papageorgi, I., Haddon, E., Creech, A., Morton, F., Bezenac, C., Himonides, E., Potter, J., Duffy, C., Whyton, T., & Welch, G. (2010). Institutional culture and learning II: inter-relationships between perceptions of the learning environment and undergraduate musicians' attitudes to performance. *Music Education Research, 12*(4), 427-446. doi: 10.1080/14613808.2010.520432
- Papageorgi, I., Creech, A. & Welch, G. (2013). Perceived performance anxiety in advanced musicians specializing in different musical genres. *Psychology of Music, 41*(1), 18-41. doi: 10.1177/0305735611408995
- Randles, C. (2011). "What is a good musician?" An analysis of student beliefs. *Arts Education Policy Review, 112*(1), 1-8. doi: 10.1080/10632913.2010.490774
- Ritchie, L., & Williamon, A. (2007). Measuring self-efficacy in music. In A. Williamon, & D. Coimbra (Ed.), *Proceedings of the International Symposium on Performance Science 2007* (pp. 307-312). Utrecht, The Netherlands: Association Européenne des Conservatoires (AEC).
- Ritchie, L., & Williamon, A. (2011a). Measuring distinct types of musical self-efficacy. *Psychology of Music, 39*(3), 328-344. doi: 10.1177/0305735610374895
- Ritchie, L., & Williamon, A. (2011b). Primary School Children's Self-Efficacy for Music Learning. *Journal of Research in Music Education, 59*(2), 146-161. doi: 10.1177/0022429411405214
- Ritchie, L., & Williamon, A. (2013). Measuring musical self-regulation: Linking processes, skills, and beliefs. *Journal of Education and Training Studies, 1*(1), 106-117. doi: 10.11114/jets.v1i1.81
- Sherer, M., Maddux, J. E., Mercandante, B., Prentice-Dunn, S., Jacobs, B., & Rogers, R. W. (1982). The self-efficacy scale: Construction and validation. *Psychological Reports, 51*(2), 663-671. doi: 10.2466/pr0.1982.51.2.663
- Upitis, R., Abrami, P. C., Varela, W., King, M., & Brook, J. (2017). Student experiences with studio instruction. *Music Education Research, 19*(4), 410-437. doi: 10.1080/14613808.2016.1202221
- Urruzola, M. V., & Bernaras, E. (2020). Music performance anxiety in 8- to 12- Year-Old children. *Revista de Psicodidáctica (English ed.), 25*(1), 76-83. doi: 10.1016/j.psicoe.2019.10.003
- Varela, W., Abrami, P. C., & Upitis, R. (2016). Self-regulation and music learning: A systematic review. *Psychology of Music, 44*(1), 55-74. doi: 10.1177/0305735614554639

Watson, K. (2010). The effects of aural versus notated instructional materials on achievement and self-efficacy in jazz improvisation. *Journal of Research in Music Education*, 58(3), 240-259. doi: 10.1177/0022429410377115

Wehr-Flowers, E. (2006). Differences between male and female students' confidence, anxiety, and attitude toward learning jazz improvisation. *Journal of Research in Music Education*, 54(4), 337-349. doi: 10.1177/002242940605400406

Zarza, F. J., Casanova O., McPherson, G. E., & Orejudo, S. (2020). Music Self-Efficacy for Performance: An Explanatory Model Based on Social Support. *Frontiers in Psychology*, 11, 1249. doi: 10.3389/fpsyg.2020.01249

Zarza, F. J., Orejudo, S., Casanova, O., & Mazas, B. (2016). Kenny Music Performance Anxiety Inventory: Confirmatory factor analysis of the Spanish version. *Psychology of Music*, 44(3), 340-352. doi: 10.1177/0305735614567932

Zelenak, M. S. (2010). Development and Validation of the Music Performance Self-Efficacy Scale. *MERI. Music Education Research International*, 4, 31-43. <http://cmer.arts.usf.edu/content/articlefiles/3122-MERI04pp31-43.pdf>

Zimmerman, B. J. (2000). Self-Efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25(1), 82-91. doi: 10.1006/ceps.1999.1016

Zubeldia, M., Díaz, M., & Goñi, E. (2018). Autoconcepto, atribuciones causales y ansiedad-rasgo del alumnado de conservatorio. Diferencias asociadas a la edad y al género. *Psychology, Society & Education*, 10(1), 79-102. doi: 10.25115/psye.v10i1.1048

Table 1. Internal consistency analysis (Cronbach's α)

Scale	Nº of items	Initial questionnaire			Nº of items	Revised questionnaire		
		Total sample	<i>Conserv. Profesional</i>	<i>Conserv. Superior</i>		Total sample	<i>Conserv. Profesional</i>	<i>Conserv. Superior</i>
Self-efficacy for learning	11	.749	.778	.680	10	.773	.792	.734
							without item 9	
Self-efficacy for performing	11	.749	.759	.721	10	.773	.773	.780
							without item 10	

Table 2. Exploratory and confirmatory factor analysis

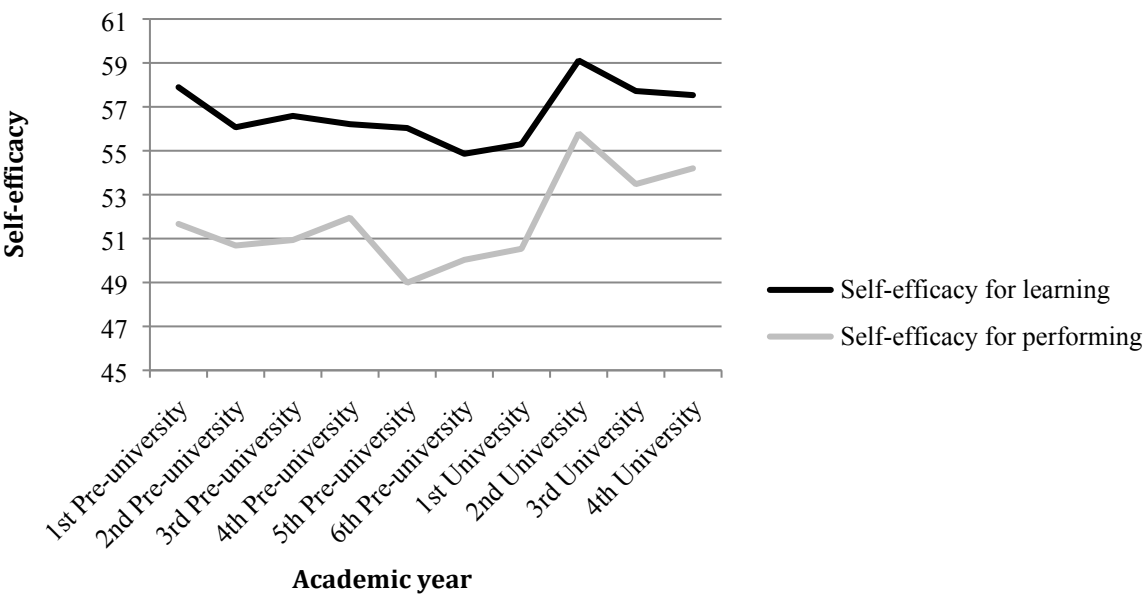
Items	EFA			CFA		CFA	
	Sub-samples 1			2 Factors		2 Factors x 2 Scales	
	1	2	3	Sub-sample 2		Sub-sample 2	
				Item	β	Item	β
Learn_1	.564			Factor 1		Factor	1
Learn_2		.407		Learn_1	.448	Learn_1	0.512
Learn_3				Learn_2	.425	Learn_3	0.529
Learn_4		.429		Learn_3	.428	Learn_6	0.685
Learn_5		.549		Learn_4	.473	Learn_7	0.559
Learn_6	.493			Learn_5	.588	Factor	2
Learn_7				Learn_6	.536	Learn_2	0.445
Learn_8		.671		Learn_7	.360	Learn_4	0.488
Learn_9			.791	Learn_8	.548	Learn_5	0.604
Learn_10		.694		Learn_10	.632	Learn_8	0.571
Learn_11		.464		Learn_11	.510	Learn_10	0.757
						Learn_11	0.598
Perf_1	.754			Factor 2		Factor1	
Perf_2		.410		Perf_1	.511	Perf_1	0.658
Perf_3		.454		Perf_2	.469	Perf_5	0.592
Perf_4		.438		Perf_3	.391	Perf_9	0.637
Perf_5	.631			Perf_4	.559	Perf_11	0.599
Perf_6		.501		Perf_5	.549	Factor 2	
Perf_7	.537			Perf_6	.580	Perf_2	0.416
Perf_8	.478			Perf_7	.637	Perf_3	0.398
Perf_9	.610			Perf_8	.656	Perf_4	0.469
Perf_10			.812	Perf_9	.456	Perf_6	0.533
Perf_11	.700			Perf_11	.546	Perf_7	0.633
						Perf_8	0.677

Table 3. ANOVA according to music education level, sex and age group

Variable		N	Mean	SD	F	Sig.	η^2
Self-efficacy for learning							
Music education level	<i>Profesional</i> (music school)	443	56.20	8.31	2.148*	.144	.003
	<i>Superior</i> (univ. level)	189	57.16	7.27			
	Total	632	56.49	8.02			
Sex	Boys	290	56.57	7.97	.017	.896	.000
	Girls	326	56.66	7.96			
	Total	616	56.62	7.96			
Age	<16	180	57.33	8.272	3.456	.016	.016
	16-17	142	54.73	7.609			
	18-20	190	56.44	7.910			
	>20	120	57.37	8.048			
	Total	632	56.49	8.022			
Self-efficacy for performing							
Music education level	<i>Profesional</i> (music school)	451	50.85	9.391	7.245	.007	.011
	<i>Superior</i> (univ. level)	186	52.99	8.523			
	Total	637	51.48	9.192			
Sex	Boys	293	53.01	8.709	14.473	.000	.023
	Girls	328	50.23	9.391			
	Total	621	51.54	9.174			
Age	<16	192	52.18	9.818	3.100	.026	.014
	16-17	135	49.50	8.503			
	18-20	188	51.44	9.064			
	>20	122	52.61	8.857			
	Total	637	51.48	9.192			

*F Based on Brown-Forsythe

Figure 1. Self-efficacy by academic year



Supplementary Material

Procedure

The translation, adaptation, and validation procedure was carried out according to instructions contained in Carretero-Dios and Pérez (2005) and Ramada-Rodilla et al. (2013). After an initial direct translation by the research team to ensure equivalency in terms of concepts, semantics, and content, we initiated a twofold procedure. On the one hand, a first pilot test was carried out on a sample of 61 music students. On the other hand, we asked two professional translators (a bilingual American musician/translator and a British psychologist specialized in translation) to produce an inverse translation, a process which consists in retranslating the questionnaire back into English and subsequently validating it. Both professionals were thoroughly satisfied with the similarity among the versions, thereby confirming that a satisfactory inverse translation procedure had taken place – apart from two items. Discrepancies on items 8 and 11 in the “self-efficacy for performance” scale were due to a difficulty of semantic translation from English to Spanish; to correct itself, the research team introduced an explanatory nuance that was not in the original. Finally, taking the two professional translators’ suggestions into account, we obtained the definitive version of each scale, incorporating slight semantic modifications: *confío-estoy seguro, los pasajes-las partes, capaz de preparar la actuación yo solo-capaz de hacer una buena actuación, prefieren que toques tu instrumento a otras actividades extraescolares-actividades/carrera, chicos-amigos, and molabas-molas*. After having obtained this final version translated into Spanish, the scale was subjected to analytical procedures in order to determine the Spanish version’s reliability and validity.

To analyze stability over time, we carried out a second round of data collection in two of the six selected conservatories and recontacted the students who had taken part in the first round, thereby obtaining a total of 75 participants. This procedure was made possible by a student identification code made up of the two first letters of their first name and the two last names.

Statistical Procedure

Initial analysis of the items was carried out on the basis of mean and standard deviation, as well as asymmetry and kurtosis indices. We likewise obtained an internal consistency measure with Cronbach’s alpha and the corrected item-total correlation, a good preliminary option prior to further analysis (Lloret-Segura et al., 2014). To obtain the factorial structure of the entire battery of questionnaires, we carried out an exploratory procedure (EFA, $n = 336$) and a confirmatory analysis (CFA, $n = 332$) of two independently and randomly generated subsamples. Although current alternatives to both options exist (such as ESEM factor analysis), our sample was sufficiently large to allow us to divide it at random and apply the traditional validation procedure (Byrne, 2012; Lloret-Segura et al., 2014). This initial portion of analysis was carried out with the SPSS statistical software in its 22.00 version, a program package that offers the usual statistical procedures in treatment of quantitative data. It uses a variance-covariance matrix, and which is appropriate for Likert-type scales featuring less than seven answer options, as here.

To apply confirmatory factor analysis (CFA) to the second subsample, we used the M-plus program, which allows for the estimation of maximum likelihood with

robust standard errors, and which is not affected by non-compliance with the assumption of multivariate normal distribution. In a final step, in order to apply factorial invariance analysis according to the recommendations of Byrne (2012), we first established a model in each of the groups, after which we adjusted for the total sample. Following Elosua (2005), the levels of invariance we took into account were configural, metric, and strict factorial invariance.

Analysis of temporal stability (test-retest) provided relevant information regarding the reliability of both scales in their Spanish versions.

Following the recommendations given by Byrne (2012), we carried out a factorial invariance study on the groups defined by the “academic level” variable (pre-university or university level). After having determined the model in each of the groups, we proceeded to general adjustment. The invariance levels taken into account were configural, metrical, and strict invariance (Elosua, 2005).

Finally, in order to complete the validation process, we analyzed differences among means (ANOVA) in each one of the factors of musical self-efficacy: regarding the groups defined by socio-demographic and pedagogical variables included in the questionnaire dossier (academic level, gender, current age and instrument).

Factorial invariance

With the purpose of establishing factorial invariance between the two academic levels featured in this study (pre-university and university-level), and following the recommendations of Byrne (2012), we started by adjusting the model independently in each of the two groups. The results of that adjustment are displayed in Table 1. It is notable that this model only maintains the three variances among errors of the global model in the subgroup of university-level students. This is to be expected, since the group of university-level students is larger. We then tested models for the three analyzed levels of invariance: configural, metric, and strict. As shown by the model adjustments, configural variance can be assumed (Table 4). In other words, in both samples the items are distributed according to the same factors; factorial weights are nevertheless not the same in the two groups, since in the more well-adjusted model only partial yet incomplete metric invariance is assumed. Three items do not reach the same regression weights: they are Items 1 and 4 in self-efficacy for learning, and Item 5 in self-efficacy for performance. Estimates on these items yield lower regression weights in Item 1 of learning and Item 5 of performance in the subsample of pre-university students when compared with the subsample of university-level students ($\beta_{Learn_1} = .217$ vs. $\beta_{Learn_1} = .518$; $\beta_{Perf_5} = .123$ vs. $\beta_{Perf_5} = .517$). On the other hand, the situation is inverted for Item 4 in learning, where university-level students are less closely related with the rest ($\beta_{Learn_4} = .399$) than pre-university students ($\beta_{Learn_4} = .601$).

Table 4. Goodness of fit indices: CFA models and Invariance

Model	Description	χ^2	DF	Sig.	χ^2/DF	RMSEA	CFI	TLI	Akaike
Based on AFE									
1	22 items	460.434	139	< .001	3.38	.072	.765	.736	23335.688
2	20 items	381.637	166	< .001	2.29	.062	.826	.801	23244.605
Based in original scale									
3	2 factors	84.908	34	< .001	2.49	.048	.939	.919	21383.887
4	1 factor	209.135	35	< .001	5.97	.088	.792	.732	21545.249
5	2 factors	226.979	34	< .001	6.67	.094	.802	.738	23314.955
6	1 factor	322.536	35	< .001	9.21	.113	.705	.621	23454.111
7	4 factors	509.184	164	< .001	3.10	.057	.856	.833	44210.058
Measurement invariance									
<i>Profesional</i> (3 error variances)		487.877	166	< .001	2.939	.066	.812	.785	31552.453
<i>Superior</i>		457.927	169	< .001	2.710	.095	.647	.603	12585.139
Configural invariance		1035.869	335	< .001	3.092	.084	.740	.705	40162.139
Metric invariance		1067.454	353	< .001	3.024	.083	.735	.714	40177.553
Partial metric invariance (3 items)		1038.084	349	< .001	2.974	.082	.744	.721	40148.655
Strict		1268.833	373	< .001	3.402	.090	.667	.661	40364.400

References

- Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic Concepts, applications, and programming*. Routledge/Taylor & Francis Group.
- Carretero-Dios, H., & Pérez, C. (2005). Normas para el desarrollo y revisión de estudios instrumentales. *International Journal of Clinical and Health Psychology*, 5(3), 521-551. <https://www.redalyc.org/pdf/337/33705307.pdf>
- Elosua, P. (2005). Evaluación progresiva de la invarianza factorial entre las versiones original y adaptada de una escala de autoconcepto [Progressive way in strict factorial invariance]. *Psicothema*, 17(2), 356-362. <http://www.psicothema.com/pi?pii=3112>
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). Exploratory Item Factor Analysis: A practical guide revised and updated. *Anales de Psicología / Annals of Psychology*, 30(3), 1151-1169. doi: 10.6018/analesps.30.3.199361
- Ramada-Rodilla, J. M., Serra-Pujadas, C., & Delclós-Clanchet, G. L. (2013). Adaptación cultural y validación de cuestionarios de salud: revisión y recomendaciones metodológicas [Cross-cultural adaptation and health questionnaires validation: revision and methodological recommendations]. *Salud Pública de México*, 55(1), 57-66. doi: 10.1590/s0036-36342013000100009

Versión española *General Musical Self-efficacy Scale* de Ritchie y Williamon (2007, 2011). Cuartero, Zarza, Casanova and Orejudo (2022).

Nos gustaría que pensaras en un actuación reciente en la que has tenido una participación importante (por ejemplo un concierto como solista de una sonata, un concierto de ensemble sobre una pieza conocida de música de cámara, un concierto que requiera improvisación sobre una melodía estándar, etc.).

Ahora, imagina que te han propuesto realizar una actividad similar en las próximas semanas (es decir, programa con nivel técnico y musical similar, interpretado en un contexto similar, con el mismo nivel de expectativas y demandas, etc.). Por favor, indica el grado de acuerdo o desacuerdo con cada uno de los siguientes enunciados, específicamente los referidos al aprendizaje y preparación para esta actuación.

AUTOEFICACIA PARA LA INTERPRETACIÓN MUSICAL		0%100%						
1	Estoy convencido de que puedo hacer la actuación satisfactoriamente.	1	2	3	4	5	6	7
2	Me he marcado unas metas altas en esta actuación, pero difícilmente las podré conseguir.	1	2	3	4	5	6	7
3	Es probable que evite o modifique los pasajes más comprometidos de la actuación.	1	2	3	4	5	6	7
4	Si percibo que el contexto alrededor de esta actuación es muy estresante (público, sala, tiempo de preparación, repertorio), intento evitarla.	1	2	3	4	5	6	7
5	Si ocurre algo inesperado durante la actuación, puedo conllevarlo bien.	1	2	3	4	5	6	7
6	Si el programa es muy difícil para mí, probablemente intente evitar la actuación.	1	2	3	4	5	6	7
7	Me siento inseguro respecto a la interpretación para esta actuación.	1	2	3	4	5	6	7
8	Es probable que si tengo una dificultad durante la actuación no sea capaz de superarla.	1	2	3	4	5	6	7
9	Soy capaz de superar los problemas que pueden aparecer durante la actuación.	1	2	3	4	5	6	7
10	La perspectiva de fracasar en esta actuación me hace trabajar más duro.	1	2	3	4	5	6	7
11	Soy capaz de hacer una buena actuación.	1	2	3	4	5	6	7