

Teachers' reasons for using self-assessment: A survey self-report of Spanish teachers.

Ernesto Panadero ¹, Gavin Brown ² & Matthew Courtney ²

Authors note

¹ Department of Educational Sciences and Teacher Education. Learning and Educational Technology Research Unit (LET), University of Oulu, Finland.

² School of Learning, Development & Professional Practice, Faculty of Education, The University of Auckland, New Zealand.

Recommended citation:

Panadero, E., Brown, G. T. L., & Courtney, M. G. R. (2014). Teachers' reasons for using self-assessment: A survey self-report of Spanish teachers. *Assessment in Education: Principles, Policy & Practice*, 21(3), 365-383. doi:10.1080/0969594X.2014.919247

This is a pre-print of an article published in *Assessment in Education: Principles, Policy & Practice*. Personal use is permitted, but it cannot be uploaded in an Open Source repository. The permission from the publisher must be obtained for any other commercial purpose. This article may not exactly replicate the published version due to editorial changes and/or formatting and corrections during the final stage of publication. Interested readers are advised to consult the official published version.

Correspondence concerning this article should be addressed to: Ernesto Panadero, Learning and Educational Technology Research Unit (LET), University of Oulu, Finland.

E-mail: ernesto.panadero@oulu.fi

Acknowledgements: Research funded through grant to Ernesto Panadero by the Alianza 4 Universidades. Support for the data analysis was given by The University of Auckland Faculty Research Development Fund grant #3701227.

Abstract

The study aimed to explore teachers' beliefs about student self-assessment and their self-reported use of self-assessment in their classrooms. A representative sample of 944 Spanish teachers (38.8% primary school sector, 54.0% secondary school sector) and 7.2% university or adult education sector) were surveyed about student self-assessment (SSA). Data were analysed using confirmatory factor analysis and structural equation modeling to determine statistically and theoretically significant predictors of teachers' self-reported use of SSA. Results showed that 90% of the teachers indicated having used SSA in their courses, 90% reported having positive experiences. A well-fitting structural model found five statistically significant predictors for the use of SSA: (1) positive experience with self-assessment, (2) belief in student participation in assessment, (3) willingness to include self-assessment as percentage of final grade, (4) self-assessment advantages, and (5) having attended assessment courses. Statistically significant mean score differences for these factors were found according to level of employment and level of educational qualifications. In conclusion, teachers' values, attitudes, and prior experiences with this type of assessment contributed a significant proportion of self-reported SSA usage.

KEYWORDS: Student self-assessment, teacher beliefs, Spain, structural equation modeling.

Notes on contributors

Ernesto Panadero received his Ph.D. degree in educational psychology from the Universidad Autónoma de Madrid (Spain) in 2011. He has worked as a post-doctoral researcher for one year at the Universitat Autònoma de Barcelona (Spain) and it is currently working at the Department of Educational Sciences and Teacher Education within the Learning and Educational Technology Research Unit (LET), at the University of Oulu (Finland). His research covers aspect related to self-regulation and formative assessment.

Gavin Brown is an Associate Professor in the School of Learning, Development & Professional Practice and Director of the Faculty Quantitative Data Analysis and Research Unit. He is a psychometrician and has carried out research into the social psychology of teacher and student beliefs about assessment. He has been involved in comparative studies with colleagues in Cyprus, Spain, Queensland, Hong Kong, and China. Previously he was a test developer for NZCER and the asTTle project at the University of Auckland, and worked for 3 years at the Hong Kong Institute of Education.

Matthew Courtney is a Ph.D. candidate at the School of Learning, Development & Professional Practice. He is a trained secondary school teacher with tertiary teaching experience. His dissertation title is: “Emerging Social Space: Social Network Engagement, Integration and Re-Enrolment of Tertiary International Students in Australasia”. His research interests are International Higher Education, Media Studies, Student Integration and Retention. He has ample experience in quantitative methodology and data analysis.

Teachers' reasons for using self-assessment: A survey self-report of Spanish teachers

1. Introduction

What teachers believe (i.e., their conceptions) makes a difference to the pedagogical strategies they might use in their classrooms (Fives & Buehl, 2012; Thompson, 1992). This is also the case for aspects related to assessment techniques and practices (Cizek, Fitzgerald, & Rachor, 1995; Kahn, 2000). If there is one assessment technique that seems most troubling to teachers it is student self-assessment (SSA); many different conceptions about the use of student self-assessment and its implementation in classroom settings exist (Tan, 2012). The purpose of this paper is to explore the beliefs teachers have about student self-assessment and how these beliefs influence their self-reported uses of SSA in their classrooms.

Student self-assessment has received considerable attention in empirical research as indicate in reviews (Boud & Falchikov, 1989; Brown & Harris, 2013; Dochy, Segers, & Sluijsmans, 1999; Falchikov & Boud, 1989; Tan, 2012). SSA is one of the key elements of formative assessment (a.k.a. assessment *for* learning) since its use leads to the development of self-regulation (Panadero, Alonso-Tapia, & Huertas, 2012) and can enhance learning and achievement (Brown & Harris, 2013; Nicol & McFarlane-Dick, 2006), giving reason to use student self-assessment in the classroom. However, what it has not been studied in such depth is the degree of SSA implementation in classrooms across educational levels. Less well understood are the reasons teachers have for using SSA or not, and whether issues of accuracy in SSA are a major factor affecting its implementation.

1.1 Advantages of the use of self-assessment

Self-assessment is about judging, evaluating, and considering the qualities of one's own academic work or abilities. In general, it seems that self-assessment practices can be grouped into three major types: self-ratings, self-estimates of performance on formal assessments, and criteria or rubric based assessments (Brown & Harris, 2013). Self-rating requires students to

judge quality or quantity aspects of their work using a rating system (e.g., a checklist or traffic lights) (Black & Harrison, 2001; Clarke, 2005). Self-estimates of performance include marking or grading one's own work using either a marking guide for objectively answered questions or a model answer (Todd, 2002). Rubrics arrange quality indicators in incremental progressions that students use to best fit the various aspects of their work and are especially common in writing or portfolio assessment (Andrade & Valcheva, 2009).

In a recent review of student self-assessment in K-12, Brown and Harris (2013) reported that self-assessment generally had a positive impact on academic performance (median effect size lay between $d=.40-.45$). Greater effects were seen in self-assessments that involved deep engagement with the processes affiliated with self-regulation (i.e., goal setting, self-monitoring, and evaluation against valid, objective standards). Training in diverse self-assessment strategies has been shown to lead to learning gains (Brown & Harris, 2013; McDonald & Boud, 2003; Panadero et al., 2012; Ramdass & Zimmerman, 2008; Ross, Hogaboam-Gray, & Rolheiser, 2002). Using models, answers, or teacher feedback to guide self-assessment judgments also generally improved performance (Hewitt, 2001; Olina & Sullivan, 2002). Children who contribute to the development of evaluative criteria and who subsequently use the criteria to self-assess also achieve better (Andrade, Du, & Mycek, 2010; Ross, 2006; Sadler & Good, 2006). Hence, it appears that self-assessment of a task or quality of one's own work will generally improve academic performance across a range of grade levels and subject areas, although the extent of these gains varies across studies.

An important consequence of self-assessment is that it contributes to increased self-regulation of learning (Klenowski, 1995; Ramdass & Zimmerman, 2008). Self-regulation of learning requires the exercise of meta-cognitive functioning in which the student monitors and evaluates his or her own performance and generates feedback as to what should be done next (Butler & Winne, 1995; Zimmerman, 2008). It should be no surprise then, that self-

assessment, especially when using rubrics and scripts, has been found to improve student self-regulation (Brown & Harris, 2013), including higher education students (Panadero & Alonso-Tapia, 2013a; Panadero, Alonso-Tapia, & Reche, 2013; Panadero & Romero, 2014). Nonetheless, teachers need to implement self-assessment in an appropriate manner, including giving training and practice in the process (Brown & Harris, 2013; Goodrich Andrade, 1996, Andrade & Valtcheva, 2009; Ross, 2006). Teachers need to provide space in the curriculum for SSA to happen and support students in engaging in the self-regulatory and metacognitive processes required in SSA. Because students tend to adapt to their teachers assessment practices (Andrade & Du, 2005; Cowie, 2005), it seems important to investigate reasons underlying teachers' uses (or not) of student self-assessment.

1.2 A crucial aspect: Accuracy

An important concern about SSA is whether it is accurate or not (Brown & Harris, 2013). Studies reviewed by Ross (2006) indicated that the student can be highly consistent in self-evaluations but much less reliable when compared to other measures (e.g., test scores, teacher ratings, or peer ratings). In general, in K-12 studies (Brown & Harris, 2013), the correlations between (a) student self-ratings and teacher ratings, (b) student self-estimates of performance and actual test scores, and (c) student and teacher rubric-based judgments were positive, but few studies had correlations greater than .60. It is important to note that these values were not consistent across student experience and academic proficiency. According to Brown and Harris (2013), younger children tended to be more optimistic, lenient, or generous than older students in their self-estimations of performance; while, older students' self-ratings were lower than younger students. Nevertheless, older students' self-assessments tended to correlate more strongly with teacher ratings or test scores. Additionally, higher performing students evaluated their own work more accurately and seemed to be more severe in assessing their work than their teachers; while lower ability students tended to be more lenient and less

accurate. More recently Boud, Lawson and Thompson (2013, 2014) reported, similarly, that high achievers underestimated their performance and low achievers overestimated, while average performing students were more accurate. Interestingly, the average students they studied had the most noticeable improvement in their performance, while more modest gains were found among the high and low achievers.

The difficulty of the task being learned interacts with students' ability to self-assess, with familiar and predictable tasks associated with more accurate student self-assessment. Self-assessments that use more specific, concrete standards or reference points, rather than subjective criteria (e.g., "I made an effort", "I'm good at this"), are associated with greater accuracy. Furthermore, classroom observations in the use of self-assessment in New Zealand schools (Harris & Brown, 2013) suggested that students may deliberately lie in their self-evaluations when they take place in public spaces (e.g., traffic lights) in order to protect their self-worth or value. Reliance on subjective criteria (e.g., "I made an effort", "I'm good at this") tended to be associated with lower accuracy in self-evaluations.

Consequently, there is a need for instructional input and caution when implementing self-assessment with students who are likely to be relatively inaccurate (i.e., younger or less academically proficient students). All self-assessment techniques seem to have similar ranges of agreement with external measures, and rubric-based self-assessment studies appear most promising because of the relatively high learning effects shown when students use them. The studies reviewed also pointed to the importance of reducing the subjectivity in the criteria students use to evaluate their work. The provision of rubrics and a focus on what others would deem as quality appear to be necessary for high-quality self-assessment (Andrade, 2010; Jonsson & Svingby, 2007; Panadero & Jonsson, 2013). Unsurprisingly, concern must be expressed about the wisdom of using student self-assessments as part of course grades or final summary evaluations because this introduces high-stakes consequences and potentially

reduces incentives for honest, accurate self-evaluations. Nonetheless, there are sometimes institutional pressures, perhaps more so in higher education (Falchikov & Boud, 1989), to include self-assessment data as part of formal, summative grading (Boud & Falchikov, 1989). Hence, while use for grading may be ill-advised, it may be required; consequently, it makes sense to evaluate the beliefs teachers have about the grading function in an examination of their attitudes towards, beliefs about, and purposes for self-assessment.

Even with the favorable empirical evidence that self-assessment can be accurate, there are still teachers who are reluctant to implement self-assessment, even in higher education (Tan, 2012). This is a crucial aspect because if teachers do not consider students to be accurate, they will be less motivated to use it in their classrooms. There is, however, lack of data about the frequency with which teachers use self-assessment and whether its use, or lack of use, is explained by their concerns about the accuracy of self-assessment.

1.3 What previous research says about self-assessment use in real settings

There seems to be relatively little research about the use of SSA in real classroom settings. In a recent review on self-assessment (Brown & Harris, 2013), only two studies reported use in real settings: Lasonen (1995) reported that, of the sample surveyed (346 secondary students), only half had participated in self-assessment, while Hunter, Mayenga, and Gambell (2006) reported that only 58% of their sample of 4,148 Canadian secondary teachers had ever used self-assessment, while 23% made minimal use. If we consider that Brown and Harris (2013) found just 84 studies in K-12 about self-assessment, it seems plausible to suggest that the real implementation of student self-assessment in K-12 classrooms is not well understood.

Another aspect that needs more exploration is, not only the frequency with which teachers utilize self-assessment, but their beliefs and attitudes about self-assessment. Research into teacher beliefs suggests that they act to (1) frame how teachers understand policy innovations, (2) filter out aspects which they consider inappropriate, and (3) guide teachers'

responses (Fives & Buehl, 2012) and that teacher beliefs are reasonably resistant to change (Richardson & Placier, 2001). With an increasing emphasis within formative assessment reforms on student engagement in assessment practices, there is a need for teachers to implement these assessment innovations. Given the strategic role of beliefs in educational practice, any effort to understand teacher practices should consider how teachers think about the phenomenon under investigation.

Previous research on SSA usage is mainly based on in-depth studies with small sample size. For example, Harris and Brown (2013) conducted three case studies with three teachers about their conceptions and practices of formative assessment, including self-assessment. One surprising result was that even though the three teachers implemented self-assessment in their classrooms, neither the teachers nor the students reported self-assessment being about the students becoming more reliable assessors. Nevertheless, they found three threats to the validity of classroom SSA: (a) the need for psychological and interpersonal safety in disclosing self-evaluations, (b) concerns about the accuracy of self-evaluations, and (c) student apathy towards being an assessor. Tan (2012) found that higher education teachers' conceptions of self-assessment fell into three different categories; that is, (a) Teacher-driven SSA, in which the teacher makes the final decision regardless of the accuracy of SSA, (b) Program-driven SSA, in which the teacher based decisions on program criteria or requirements; and (c) Future-driven SSA, in which the teachers leaves the students fully in charge of SSA in order to communicate trust and transfer to students. Third, Similarly, Dixon, Hawe, and Parr (2011) found that the beliefs and knowledge about feedback of 20 New Zealand teachers fell into three groups; that is, (a) technician, (b) pragmatist, and (c) empowerer. Self-assessment was absent in the views of the technicians, pragmatists accepted the logic of SSA but did little of it, while the empowerers made extensive use of SSA because of their general commitment to student development of "evaluative and productive knowledge and expertise" (Dixon et al., 2011, p.

370). In addition, another key finding was that teachers' deep seated beliefs about their role and that of students in learning and assessment were influential in regard to implementation of SSA.

What these studies reveal is that, while SSA may be seen in a positive light, teachers' beliefs about the validity and integrity of SSA may be a logical extension of more dominant beliefs about the nature of teacher and student roles in classroom contexts. There also seems to be a suggestion that the policy priorities of an employment context contribute significantly to teacher attitudes; more positive views require more formative assessment policy contexts. Further, it shows that teachers' understanding of how SSA should be implemented tends to be dominated by an instrumental, technical approach in which the benefits of SSA are ignored. Needless to say, the sample size in these three studies is rather small and more information is needed on the implementation of self-assessment in real classrooms from a larger sample to have a more complete picture of reasons behind the use of SSA.

1.4 Why Spain?

Spain is an interesting context in which to study the implementation of self-assessment due to the 1990 Reform Act (LOGSE: Ley Orgánica de Ordenación General del Sistema Educativo [Organic Law for the Education System Organization]) that promoted new methodological approaches towards changing the assessment paradigm (Remesal, 2007, 2011). The previous Education Law did not address any aspects of formative assessment, on the contrary it had a strong summative assessment perspective. In the 1990 reform act the underlying paradigm in K-12 schooling is oriented towards formative assessment purposes (Remesal, 2007, 2011). Therefore, it is interesting to explore the extent to which the reform process has impacted on the occurrence and features of self-assessment in real classrooms in primary and secondary education.

On the other hand, the Spanish higher education context does not have clear guidelines about what type of assessment should be implemented. Decisions about assessment practices rely on either the individual teachers or group of teachers lecturing the same course. At the same time there is less reflection about the learning and assessment processes, as university teachers in Spain do not have to complete any specific training on these topics, unlike primary and secondary teachers. Ion and Cano (2011) found that Spanish university teachers based their assessment mainly in traditional approaches -mostly exams and written works-, and more formative assessment approaches being less reported by the teachers. Therefore, as a contrast, it is useful to include teachers who work in contexts not directly influenced by the LOGSE; that is, higher education.

1.5 Aims

Using a self-report survey instrument, the present study explores how primary, secondary, and university teachers report making use of self-assessment. Since teacher beliefs about the roles and purposes of education phenomena, including assessment, are significant predictors of practices (Fives & Buehl, 2012), data were collected concerning teachers' training in, attitude toward, and perceptions of advantages and disadvantages of self-assessment in the Spanish educational context. Since individual beliefs and social norms are important antecedents of intention and action (Ajzen, 2005), it was decided to use structural equation modeling (SEM) to determine the explanatory effect of these factors on self-reported usage of student self-assessment.

1.5.1 Research questions

RQ1: What training in, experiences of, and values or attitudes about student self-assessment do Spanish teachers have?

RQ2: What effect do these values and experiences have on the self-reported use of self-assessment in real settings?

RQ3: How does the proposed model of self-assessment differ for subgroups of interest (e.g., Educational level)?

2. Method

2.1 Participants

A total of 1312 primary education centers and 814 secondary education centers were invited to participate in the study, constituting 11.44% of all Spanish education centres. In contrast, just seven (9.21%) of the 76 public universities in Spain were invited. The study made use of 944 teaching professionals of which a majority 703 were female (74.5%) and 129 were male (13.7%), with 112 respondents (11.9%) not giving their sex. Of the 944 respondents, 366 (38.8%) were drawn from the primary school sector, whilst 510 (54.0%) were drawn from the secondary school sector, and only 68 (7.2%) were drawn from the university or adult education sector. Given the average of 33.61 teachers per K-12 center¹, the maximum possible respondents, if all centers had chosen to participate, would be 71,453; the current study sample of 876 K-12 teachers has a margin of error of just 3.29%.

Unfortunately, the small sample of higher education instructors relative to the estimated population of instructors in seven universities (i.e., 68 / 10824) has a margin of error of 11.85%.

In terms of qualifications, 215 teachers (22.8%) had acquired technical certificates, 634 teachers (67.2%) had graduated with a bachelor's degrees, 32 (3.4%) had attained a masters, and 62 (6.6%) had attained a PhD. While relatively small in number, the higher education instructors form a naturally occurring contrast group for the K-12 teachers who are under the jurisdiction of the LOGSE framework on educational assessment. Where feasible,

¹ As published by the Spanish Educational Department in in 2014 (<https://www.mecd.gob.es/servicios-al-ciudadano-mecd/estadisticas/educacion/indicadores-publicaciones-sintesis/cifras-educacion-espana/2014.html>)

contrasts will be made between teachers working in the different levels of the education system so as to explore the relationship of teacher self-reported attitudes and beliefs conditioned upon their employment circumstances.

Our sample can be considered as a convenience one as it was based on the commitment of the centre leaders to distribute the information and link for the survey to the teachers in the centre and participation was voluntary. The respondents were drawn from various regions across Spain (Table 1), with only two regions having more than 100 participants (i.e., Andalucía and Madrid). While the sample is not representative of the higher education population, the sample is large enough for K-12 teachers to allow reasonable generalisations about Spanish teacher beliefs in the compulsory schooling sector.

<table 1 about here>

2.2 Instrument

A self-report survey instrument including a total 14 questions concerning self-assessment (SSA) was administered in 2012. The questionnaire sought teachers' demographic information and relevant employment data (i.e., sex, level of employment, level of qualifications, level of assessment training, and location in Spain). Questions were also posed concerning teachers' use of SSA in their courses or for grading purposes, their experience with SSA, their opinions as to the advantages and problems of SSA, including its accuracy. A full list of questions and possible response options is available in Appendix A.

2.3. Procedure

Educational centres were contacted via phone ($n = 677$) or via email ($n = 1456$). Requests for participation were sent to the principal of each education centre who was asked to refer the request to complete an online survey to the teachers in the centre. A total of 1312 primary education centres, 814 secondary education centres, and seven public universities

were contacted. Centres in all Spanish regions were contacted for primary and secondary education so as to generate a representative sample of Spanish K-12 schooling.

2.4 Analysis

To answer RQ1 concerning the extent of training in assessment, use and overall impression of self-assessment, basic descriptive statistics (i.e., mean, standard deviation, and frequency) are utilized. Tabulated results are reported for primary, secondary, and university/adult levels of employment for the purpose of comparison. Since there were four items exploring the pros and cons of SSA each, a multiple indicator, latent trait theory approach was used to simplify the dimensionality of the items. A two-factor solution (i.e., advantages and disadvantages) was tested for quality of fit to the data and admissibility with confirmatory factor analysis (CFA).

Since the goal of RQ2 was to understand how teacher beliefs and experiences contributed to self-reported usage of SSA, a causal-correlational approach was used to identify the relationship of latent and manifest variables to each other and their contribution, if any, to SSA usage. Stepwise regression is well suited to identifying the unique contribution of multiple simultaneous predictors to a construct of interest (Menard, 1995); this is achieved by introducing conceptually related groups of predictors sequentially into a model until no gain in variance explained (R^2) is achieved. Structural equation modelling (SEM), based on these exploratory analyses, was used to evaluate the quality of the model fit to the data. SEM allows for the inclusion of both observed variables and latent factors in one model, as well as permitting multiple inter-correlated predictors of a construct, providing a more sophisticated evaluation of the relationship of multiple predictors to each other and to a dependent construct (Bollen, 1989; Borsboom, 2006).

To answer RQ3 about differences in the model according to teacher demographic factors, a multiple-indicator, multiple-cause (MIMIC) approach for determining differences

in latent factor and observed variable means was utilized. This approach allows the identification of the unique effect of each demographic characteristic upon the mean score for each component of the model more accurately than the conventional analysis of variance approach (Hancock, Lawrence, & Nevitt, 2000). The MIMIC approach was used because the small sample sizes resulted in some missing cells in bivariate tables, making the alternative structured means modeling approach (SMM, Sörbom, 1974) unsuitable. To do this, two dummy variables for teacher education level (i.e., three levels—primary, secondary, tertiary), teacher qualification level (i.e., technical certificate/ vocational training technical discipline, bachelor degree, masters degree, and doctorate), were created. Teaching experience was used as a continuous variable according to the number of years teaching experience each participant reported. These demographic factors were regressed, separately in three analyses, as a causal indicator on the factor and six manifest variables in the SEM model.

Because the questionnaire used binary, ordinal, and continuous variables, the WLSMV estimator in Mplus was utilized for all EFA, CFA, and SEM procedures (Muthén & Muthén, 1998-2010). This estimator handles appropriately two- and three-option response scales (Beauducel & Herzberg, 2006). The polychoric/tetrachoric correlation coefficient between the proposed independent variables and dependent variable, for the purpose of the stepwise procedure, was also estimated individually using Mplus (2010) WLSMV estimator. Model fit in CFA and SEM is determined by inspection of multiple indices (Hu & Bentler, 1999; Fan & Sivo, 2005); conventional standards for good fit are statistically non-significant probability ($p > .05$) for the ratio of χ^2/df (Marsh, Hau, & Wen, 2004), comparative fit index (CFI) and gamma hat ($\hat{\gamma}$) $> .95$ (Fan & Sivo, 2007); root mean square error of approximation (RMSEA) < 0.05 (Hu & Bentler, 1999), and the weighted root mean residual (WRMR) < 1.00 (Muthén & Muthén, 2010).

3. Results

3.1 Student Self-assessment from the Teacher Perspective

RQ1 is answered by a number of variables. Most teachers have completed training in assessment courses (Table 2), with greater completion (approximately 80%) by primary and secondary teachers, compared to only half of the university adult education teachers.

<table 2 about here>

The vast majority of teachers (83.5%) agreed that students ought to participate in assessment (Table 3), with similar high endorsement among primary and secondary teachers, relative to higher education instructors. The majority of teachers (62.3%) were willing to include self-assessment as a percentage of their final course grade or were already doing it (Table 3). Proportions in each category were very similar across the three teaching level. Almost all teachers (93.6%) indicated that they already used self-assessment in their current courses (Table 3), though small decreases in the proportion reporting the use of self-assessment were observed as teaching sector increased.

<table 3 about here>

Most teachers (87.4%) reported having a positive experience of student self-assessment (Table 4), though as teaching level increased, the positive experience rating declined and the neutral response increased. Over half the teachers indicated that they thought student self-assessment was not accurate (56.8%), with just a quarter (27.3%) indicating that student self-assessments were accurate. Generally, compared to primary and secondary teachers, higher education teachers considered student self-assessment as a more accurate assessment tool.

<table 4 about here>

Teachers were asked to indicate whether they agreed with four possible advantages or disadvantages of self-assessment. Advantages included perceived improvement in (a) student responsibility for learning, (b) detection and correction of problems, (c) time saving for

teachers, and (d) student learning with the use of this strategy. Disadvantages included the strategy's (a) unreliability, (b) threat to teacher authority, (c) waste of time, and (d) lack of benefit to student learning. In contrast to expectations, confirmatory factor analysis identified just one admissible, well-fitting latent construct (i.e., Advantages of Self-Assessment) consisting of three advantage items (i.e., Detection and Correction of Problems $\beta = .62$; Saves Time for the Teachers $\beta = .71$; and Students Learn Using This Strategy $\beta = .71$). The remaining items had low loadings or strong cross-loadings and, so were dropped from further analysis.

3.2 Contributors to the Use of Self-Assessment

To answer RQ2 about the significant contributors to the reported use of self-assessment in real settings, a stepwise regression modeling approach was used. It was presumed that the instructors' responses to five constructs (i.e., beliefs about the advantages of SSA, positive experience of SSA, participation in assessment training, willingness to let students participate in assessment, and willingness to include SSA in final grades) would predict their self-reported usage of SSA. Further, it was presumed that these predictors themselves would be relatively independent of each other, though a degree of inter-correlation was possible. Except for the latent trait Advantages of Self-Assessment, all other relevant variables were used in manifest form (Table 5).

<table 5 about here>

In the order presented in Table 5, each factor/item was placed into a stepwise regression model until no further significant contribution was observed ($p < 0.05$). Based on the stepwise regression result, a well-fitting ($\chi^2 = 16.08$, $df = 10$, $\chi^2/df = 1.61$, $p = 0.20$, CFI = 0.99, RMSEA = 0.03, WRMR = 0.54, $\hat{g} = 0.998$) structural equation model was created (Figure 1). The model shows that there are four single-response items and one latent factor that have statistically significant contributions to the Use of SSA item. Of these five

predictors, three (i.e., previous positive experience with SSA, advantages of SSA, and previous training in assessment) had moderate standardised regression weights ($\beta > .20$), accounting for most of the variance in Use of SSA.

<Figure 1 about here>

The inter-correlations between the five predictors (i.e., the SSA Advantage Focused factor and other four manifest indicators) were generally weak, ranging from $.01 \leq r \leq .28$, indicating that each construct provided considerable unique variance to the model. Each predictor provided a small but statistically significant contribution to Use of Self-Assessment responses ($.13 \leq \beta \leq .33$). Total variance explained for Use of Self-Assessment was $R^2 = .42$, a large effect size $f^2 = .72$ (Cohen, 1992).

3.3 An Assessment of Latent and Observed Mean Difference

Having established a general model, the effect of teacher demographic characteristics was assessed using the MIMIC analysis of means, as detailed in section 2.3. Three separate analyses were run for the difference in latent factor and observed variable means according to (a) educational level: (i.e., primary = 1, secondary = 2, and 3 = university = 3), (b) qualification level of teachers (i.e., 1 = technical certificate/ vocational training technical discipline = 1, bachelor = 2, master = 3, and PhD = 4), and (c) teaching experience (i.e., years).

Simple inspection of mean proportions might suggest that university teachers have less formal training in assessment (Table 2), have a higher proportion believing it is not necessary to include SSA (Table 3), and have a higher proportion believing SSA can be accurate (Table 4). Instead, the means analysis (Table 6) has showed that only one of these predictions was statistically significant. University teachers had a statistically significant lower (a) positive experience of SSA, (b) belief in student participation in assessment, and (c) use of self-assessment. Similarly, increased levels of teacher qualifications produced

statistically significant decreases in (a) positive experience with SSA, (b) participation in assessment training, (c) belief in student participation in assessment, and (d) use of self-assessment. Teaching experience had no statistically significant effect on mean scores. Hence, university teachers, unlike primary and secondary teachers, tended not to accept the importance and utility of SSA.

<table 6 about here>

3.4 Summary of Results

This sample of teachers, especially those in K-12, have had courses in assessment, agreed that students needed to participate in assessment, and claimed to be using SSA in their courses, with some greater reluctance to use SSA as part of a final course grade. Most have had a positive experience of SSA, but still considered SSA to be inaccurate. Nonetheless, there was a focus on the advantages, rather than obstacles, in SSA. Five predictors were found for self-reported use of self-assessment and statistically significant differences in mean scores were found for three variables according to level of employment and level of educational qualifications.

4. Discussion

This cross-sectional survey study clearly shows that self-reported use of student self-assessment in classroom settings across educational levels in Spain is strongly influenced by teachers' values, attitudes, and prior experiences with this type of assessment. The strongest predictors of use were (a) previous positive experience of SSA, (b) endorsement of the educational advantages of SSA, and (c) previous training in assessment. Weaker, but statistically significant, predictors were belief in SSA and willingness to incorporate SSA in final course grades.

The relatively larger-scale endorsement of SSA in the compulsory school sector suggests that teachers have very much complied with and, possibly, embraced the formative

requirements of the 1990 LOGSE reform. In contrast, higher education practices and teacher beliefs seem to be resolutely summative; perhaps, because the LOGSE reforms do not apply to the higher education sector. The strong self-reported use of SSA among the school sector teachers in Spain stands in contrast to the much weaker self-reported usage reported previously in Scandinavia and Canada. Presumably, the long-standing assessment for learning policy and various professional development deployed in the last 20 years has had a positive cumulative effect on teacher beliefs.

It is not surprising that previous positive experience of SSA, endorsement of the educational advantages of SSA, and previous training in assessment contributed to greater use of SSA. Prior experience and training combined with a focus on advantages generate endorsement of using SSA. While problems in the use of SSA were deliberately elicited in the survey, none of these had a statistically significant relationship to the use of SSA, suggesting that addressing teacher concerns about SSA would be a less powerful method of encouraging use of SSA. Rather, the data suggest that (a) ensuring teachers have a positive experience of using SSA in any way in their teaching and (b) emphasising the positive effects of SSA on teacher workload and student learning would be much more powerful professional development strategies. Such an approach would be consistent with the recommendations that there is greater success when professional development is embedded in school practices and contexts rather than carried out in formal educational experiences (Richardson & Placier, 2001). Nonetheless, the study reinforces the view that teachers' perceptions of instructional and assessment activities are very much influenced by prior experience and related beliefs.

The study indicates that it is not so much teachers' personal qualifications that determine differences between compulsory and higher education sectors. Rather, it seems more likely that policy the environment of higher education, which position assessment as predominantly a summative evaluation (Ion & Cano, 2011), underpins the differences.

Higher education courses tend to be one semester-long, strongly content-focused, and involve just 2-3 hours interaction per week between instructors and students. Within such a context, involving students in assessment may be challenging and the high-stakes consequences of assessment may elicit concerns about the validity of SSA.

It is worth considering that some of the difficulties in developing comparative models (i.e., by level of employment) is a partly a function of the restricted response scales. Having only two or three response options means that there is little variance in the participant responses. Future studies should consider using longer scales (preferably 5-7 options), and serious consideration should be given to using positively-packed rating scales (Lam & Klockars, 1982), since these are ideally designed for conditions when participants are inclined to be positive, which is certainly evident in the data. It is noteworthy that the covariances between the Positive experience with SSA factor and the other manifest variables were relatively weak. This may also be related to the restriction of range in the response categories (i.e., ten variables were dichotomous and four had 3 options). However, the WLSMV estimation method, using polychoric and polyserial correlations, does account for the attenuated covariances due to restricted scale length (Muthén & Muthén, 2010), suggesting that the observed correlations are robust estimations of the relationships among predictors.

Another limitation of this study is that data is self-reported and has not been contrasted with more objective data (e.g., observations in classrooms). While self-reports are an important first step to forming a rich picture of self-assessment usage, future studies should seek to distinguish between actual and espoused practices. A third concern rests with the limited sample size of higher education instructors; the comparative results between higher education and K-12 teachers should be taken cautiously and provides direction for future research. Fourthly, the use of a survey design means that it is not possible to prove the

causal paths described in Figure 1; future research needs to identify, perhaps qualitatively, the explicit causes of teacher use of SSA and track teachers longitudinally to ascertain whether changes in teacher thinking as a result of externally supported experiences of SSA result in greater and more sophisticated use of SSA independent of any intervention.

While the analysis was not able to compare the structural model by levels of employment due to estimation problems with small N in one group, the mean score differences showed the level of employment and level of qualification resulted in statistically significant differences for some key variables in this SSA survey. Positive experience with SSA, reported use of SSA, and belief in student participation in assessment were much less prevalent for higher education teachers, who were also likely to have higher qualifications. Hence, teacher attitudes and beliefs seem quite different between those working in the K-12 sector and those in higher education. It could be that this difference is a function of higher education instructors working in a context of greater focus on teacher-vetted and approved scores, compared to those in K-12 where the official policy is focused more on assessment for learning practices, including greater use of student self-assessment. Thus, there is some support for the idea made by Brown and Michaelides (2011) that teacher beliefs eventually correspond with the policy frameworks in which they work.

This study adds to our understanding of the impact of teacher beliefs and experiences on assessment practices. Simply, teachers have reasons for why they claim to use SSA. The study advances our understanding of the frequency with which SSA is supposedly being carried out—perhaps, it takes a generation before policy innovations, such as assessment for learning, can be seen to be having an effect. The study also points us clearly to leverage points in ensuring assessment for learning policies are implemented. Teachers need positive experiences, training, and a focus on the advantages, rather more than attention on the problems inherent in SSA. While normal policy recommendations in SSA are to isolate such

evaluations from summative grading, this study hints that making SSA part of course grading would lead to greater usage. Raising the stakes would motivate instructors, while potentially threatening SSA validity; however, the study makes clear that this policy course needs to be complemented with staff training and ensuring positive experiences are established.

Recommendations for proper implementation of SSA for coursework are available (Boud & Falchikov, 1989; Falchikov & Boud, 1989) and should be followed. If done so, this study indicates SSA would be used more and probably well.

References

Ajzen, I. (2005). *Attitudes, personality and behavior* (2nd ed.). New York: Open University Press.

Andrade, H. (2010). Students as the definitive source of formative assessment: Academic self-assessment and the self-regulation of learning. In H. J. Andrade & G. J. Cizek (Eds.), *Handbook of Formative assessment* (pp. 90-105). New York: Routledge.

Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation, 10*(3), 1-11.

<http://pareonline.net/getvn.asp?v=10&n=3>

Andrade, H., Du, Y., & Mycek, K. (2010). Rubric-referenced self-assessment and middle school students' writing. *Assessment in Education: Principles, Policy & Practice, 17*(2), 199-214. doi: 10.1080/09695941003696172

Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Into Practice, 48*(1), 12-19.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal, 13*(2), 186-203.

doi:10.1207/s15328007sem1302_2

- Black, P., & Harrison, C. (2001). Self- and peer-assessment and taking responsibility: The science student's role in formative assessment. *School Science Review*, 83(302), 43-49.
- Bollen, K. A. (1989). *Structural equations with latent variables*. Canada: Wiley.
- Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, 71(3), 425-440. doi: 10.1007/s11336-006-1447-6
- Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher-education: A critical analysis of findings. *Higher Education*, 18(5), 529-549. doi: 10.1007/BF00138746
- Boud, D., Lawson, R., & Thompson, D. G. (2013). Does student engagement in self-assessment calibrate their judgement over time? *Assessment & Evaluation in Higher Education*, 38(8), 941-956. doi: 10.1080/02602938.2013.769198
- Boud, D., Lawson, R., & Thompson, D. G. (accepted for publication). The calibration of student judgement through self-assessment: Disruptive effects of assessment patterns. *Higher Education Research and Development*.
- Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. McMillan (Ed.), *The SAGE handbook of research on classroom assessment* (pp. 367-393). Thousand Oaks, CA: SAGE.
- Brown, G. T. L., & Michaelides, M. (2011). Ecological rationality in teachers' conceptions of assessment across samples from Cyprus and New Zealand. *European Journal of Psychology of Education*, 26(3), 319-337. doi: 10.1007/s10212-010-0052-3.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65(3), 245-281.

- Cizek, G. J., Fitzgerald, S. M., & Rachor, R. A. (1995). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment, 3*(2), 159-179. doi: 10.1207/s15326977ea0302_3
- Clarke, S. (2005). *Formative assessment in the secondary classroom*. Abingdon, UK: Hodder Murray.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.
- Cowie, B. (2005). Pupil commentary on assessment for learning. *Curriculum Journal, 16*(2), 137-151. doi: 10.1080/09585170500135921
- Dixon, H. R., Hawe, E., & Parr, J. (2011). Enacting Assessment for Learning: The beliefs practice nexus. *Assessment in Education: Principles, Policy & Practice, 18*(4), 365-379. doi: 10.1080/0969594X.2010.526587
- Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer- and co-assessment in higher education. A review. *Studies in Higher Education, 24*(3), 331-350. doi: 10.1080/03075079912331379935
- Fan, X., & Sivo, S. A. (2005). Sensitivity of fit indexes to misspecified structural or measurement model components: Rationale of two-index strategy revisited. *Structural Equation Modeling, 12*(3), 343-367.
- Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research, 42*(3), 509-529.
- Fives, H., & Buehl, M. M. (2012). Spring cleaning for the “messy” construct of teachers' beliefs: What are they? Which have been examined? What can they tell us? In K. R. Harris, S. Graham & T. Urdan (Eds.), *APA Educational Psychology Handbook: Individual Differences and Cultural and Contextual Factors* (Vol. 2, pp. 471-499). Washington, DC: APA.

Harris, L. R., & Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education, 36*, 101-111.

doi: 10.1016/j.tate.2013.07.008

Hewitt, M. P. (2001). The effects of modeling, self-evaluation, and self-listening on junior high instrumentalists' music performance and practice attitude. *Journal of Research in Music Education, 49*(4), 307-322. doi: 10.2307/3345614

Hunter, D., Mayenga, C., & Gambell, T. (2006). Classroom assessment tools and uses: Canadian English teachers' practices for writing. *Assessing Writing, 11*(1), 42-65.

Hu, L-T, & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2*, 130-144.

Kahn, E. A. (2000). A case study of assessment in a grade 10 english course. *The Journal of Educational Research, 93*(5), 276-286. doi: 10.1080/00220670009598719

Klenowski, V. (1995). Student self-evaluation processes in student-centred teaching and learning contexts of Australia and England. *Assessment in Education, 2*(2), 145-163.

Lam, T., & Klockars, A. J. (1982). Anchor point effects on the equivalence of questionnaire items. *Journal of Educational Measurement, 19*(4), 317-322.

Lasonen, J. (1995). A case study of student self-assessment in upper secondary education: Institute for Educational Research, University of Jyväskylä, P.O. Box 35, 40351 Jyväskylä, Finland.

Marsh, H W, Hau, Kit-Tai, & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in

- overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11*(3), 320-341.
- McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education: Principles, Policy & Practice*, *10*(2), 209-220.
- Menard, S. (1995). *Applied logistic regression analysis*. Thousands Oaks, CA: SAGE.
- Muthén, L. K., & Muthén, B. O. (2010). *Mplus User's Guide (6th ed.)*. Los Angeles, CA: Muthén & Muthén.
- Nicol, D., & McFarlane-Dick, D. (2006). Formative assessment and self-regulated learning, a model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199-218. doi: 10.1080/03075070600572090
- Olina, Z., & Sullivan, H. J. (2002). Effects of classroom evaluation strategies on student achievement and attitudes. *Educational Technology, Research and Development*, *50*(3), 61-75.
- Panadero, E., & Alonso-Tapia, J. (2013a). Revisión sobre autoevaluación educativa: evidencia empírica de su implementación a través de la autocalificación sin criterios de evaluación, rúbricas y guiones. *Revista de Investigación en Educación*, *11*(2), 172-197.
- Panadero, E., & Alonso-Tapia, J. (2013b). Self-assessment: theoretical and practical connotations. When it happens, how is it acquired and what to do to develop it in our students. *Electronic Journal of Research in Educational Psychology*, *11*(2), 551-576. doi: <http://dx.doi.org/10.14204/ejrep.30.12200>
- Panadero, E., Alonso-Tapia, J., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences*, *22*(6), 806-813. doi: 10.1016/j.lindif.2012.04.007

- Panadero, E., Alonso-Tapia, J., & Reche, E. (2013). Rubrics vs. self-assessment scripts effect on self-regulation, performance and self-efficacy in pre-service teachers. *Studies in Educational Evaluation*, 39(3), 125-132. doi: 10.1016/j.stueduc.2013.04.001
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9(0), 129-144. doi: <http://dx.doi.org/10.1016/j.edurev.2013.01.002>
- Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance, self-efficacy and calibration. *Sent for publication to Assessment in Education: Principles, Policy & Practice*.
- Ramdass, D., & Zimmerman, B. J. (2008). Effects of self-correction strategy training on middle school students' self-efficacy, self-evaluation, and mathematics division learning. *Journal of Advanced Academics*, 20(1), 18-41. doi: 10.4219/jaa-2008-869
- Remesal, A. (2007). Educational reform and primary and secondary teachers' conceptions of assessment: The Spanish instance, building upon Black and Wiliam (2005). *Curriculum Journal*, 18(1), 27-38. doi: 10.1080/09585170701292133
- Remesal, A. (2011). Primary and secondary teachers' conceptions of assessment: A qualitative study. *Teaching and Teacher Education*, 27(2), 472-482. doi: <http://dx.doi.org/10.1016/j.tate.2010.09.017>
- Richarson, V., & Placier, P. (2001). Teacher change. In V. Richardson (Ed.), *Handbook of Research on Teaching (4th ed.)* (pp. 905-947). Washington, DC: AERA.
- Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment Research & Evaluation*, 11. <http://pareonline.net/getvn.asp?v=11&n=10>
- Ross, J. A., Hogaboam-Gray, A., & Rolheiser, C. (2002). Student self-evaluation in grade 5-6 mathematics effects on problem-solving achievement. *Educational Assessment*, 8(1), 43-59.

Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning.

Educational Assessment, 11(1), 1 - 31.

Tan, K. H. K. (2012). *Student self-assessment. Assessment, learning and empowerment.*

Singapore: Research Publishing.

Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In

D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning*

(pp. 127-146). New York: Macmillan.

Todd, R. W. (2002). Using Self-Assessment for Evaluation. *English Teaching Forum, 40*(1),

16-19.

Table 1. *Frequency and Percentage of Participants by Spanish Regions*

Region	Frequency	Percent
Andalucía	162	17.2
Madrid	103	10.9
Comunidad Valenciana	86	9.1
Murcia	69	7.3
Cataluña	60	6.4
Asturias	54	5.7
Castilla-Leon	51	5.4
Extremadura	48	5.1
País Vasco	39	4.1
Canarias	34	3.6
Aragón	26	2.8
Baleares	25	2.6
Navarra	24	2.5
Castilla-la Mancha	10	1.1
Cantabria	8	.8
La Rioja	4	.4
Galicia	2	.2
Missing	139	14.7

Note. $N=944$.

Table 2. *Completion of Training Course in Assessment*

Education Level	Positive		Negative	
	<i>N</i>	%	<i>N</i>	%
Primary	285	77.9	81	22.1
Secondary	415	81.4	95	18.6
University	35	51.5	33	48.5

Note. *N* = 944.

Table 3. *Teacher Uses Student Self-Assessment*

Response	<u>Primary</u>		<u>Secondary</u>		<u>University/Adult</u>	
	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>
<i>Necessity</i>						
Not Necessary	28	7.7	47	9.2	14	20.6
Sometimes/ Depends	21	5.7	42	8.2	4	5.9
Yes/ Yes I already do that	317	86.6	421	82.5	50	73.5
<i>Currently Use in my Courses</i>						
Yes	350	95.6	473	92.7	61	89.7
No	16	4.4	37	7.3	7	10.3
<i>Use as Part of Final Course Grade</i>						
No	105	28.7	190	37.3	22	32.4
Maybe/ Depends	18	4.9	20	3.9	1	1.5
Yes/ Yes I already do that	243	66.4	300	58.8	45	66.2

Table 4. Teachers' General Opinions about Student Self-Assessment

	<u>Primary</u>		<u>Secondary</u>		<u>University/Adult</u>	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
<i>General Experience</i>						
Negative	0	0	6	1.2	1	1.5
Neutral	31	8.5	63	12.4	18	26.5
Positive	335	91.5	441	86.5	49	72.1
<i>Accuracy of student self-assessment</i>						
No, they are not accurate	197	53.8	309	60.6	30	44.1
Perhaps*	71	19.4	68	13.3	11	16.2
Yes, they are accurate	98	26.8	133	26.1	27	39.7

Note. *=Perhaps includes the following responses: sometimes, depends, and once they have

learnt; *N*=944

Table 5. Correlation Coefficients between Use of Self-Assessment and other Self-assessment Items and Factors

#	Factor/ item	ρ
4	Positive Experience with Self-Assessment	.77***
3	Belief in Student Participation in Assessment	.45***
6	Willingness to Include Self-Assessment as % of Final Grade	.40***
(8ii-iv)	Self-Assessment Advantage Focused (factor)	.38***
2	Participation in Assessment Course	.34***
7iv	Main Problem: It does not Enhance of Student Learning	-.21
8i	Advantage: Students More Conscientious and Responsible for Learning	.18
7i	Main Problem: Unreliability	.15
7ii	Main Problem: Problem with Teacher Authority	.10
5	General Accuracy of Self-Assessment	.06
7iii	Main Problem: Loss of Time	-.02

Note. ρ = polychoric correlation coefficient, * $p < 0.05$, ** $p < 0.01$, *** $p < .001$.

Table 6. Estimation of Factor/ Item Mean Difference by MIMIC-Derived Tests of Equivalence

Factor/ Item	Demographic Influence		
	Educational	Qualification	Teaching
	Level (γ) [†]	Level (γ) [*]	Experience (γ) [*]
Positive Experience with SSA	-.15***	-.17***	.05
SSA Advantage Focused (factor)	.00	.06	-.08
Participation in Assessment Training	-.17	-.20*	.05
Belief in Student Participation in Assessment	-.10**	-.06*	.00
Willingness to Include SSA as % of Final Score	-.06	-.04	.06
Use of Self-Assessment	-.14*	-.14**	-.05

Note. [†] $N = 944$; ^{*} $N = 943$, ^{*} $p < 0.05$, ^{**} $p < 0.01$, ^{***} $p < .001$.

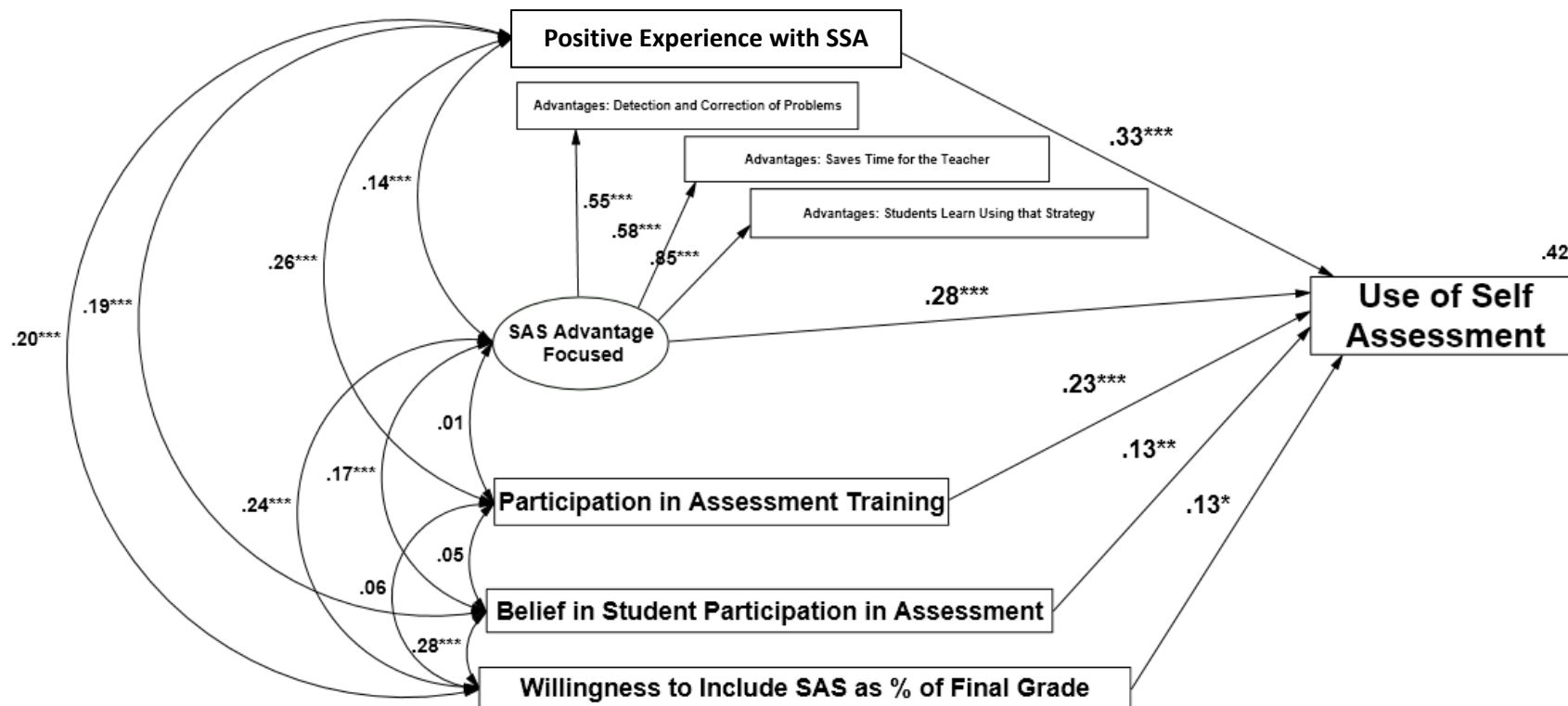


Figure 1. Schematic Model of Self-Assessment Attitudes to the Use of Self-Assessment in Spanish Educational Setting

Note. $N = 944$; * $p < 0.05$, ** $p < 0.01$, *** $p < .001$

Appendix A. English Translation of Survey Questions and Response Format

#	Question	Response format
1	Do you use self-assessment in your courses?	2-point agreement
2	Have you done any course on assessment?	2-point agreement
3	Do you think that students need to participate in the assessment?	3-point agreement
4	If you use self-assessment, how would you consider your experience?	3-point rating
5	Do you think students are accurate when self-assessing?	3-point rating 1
6	Would you let a percentage of your course grade depend on your students' self-assessment?	3-point rating 1
7i	Problem with SSA: Unreliability	2-point agreement
7ii	Problem with SSA: Creates problems with the teacher's authority	2-point agreement
7iii	Problem with SSA: It causes more loss of time than it saves	2-point agreement
7iv	Problem with SSA: It does not enhance students' learning	2-point agreement
8i	Advantages: Students are more conscious and responsible for their learning.	2-point agreement
8ii	Advantages: Detection and correction of problems	2-point agreement
8iii	Advantages: Saves time for the teacher	2-point agreement
8iv	Advantages: Students learn using that strategy	2-point agreement