**Higher Education**

**Assessing the Students' Evaluations of Educational Quality (SEEQ) questionnaire in Greek higher education**

---Manuscript Draft---

<table>
<thead>
<tr>
<th>Manuscript Number:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full Title:</strong></td>
<td>Assessing the Students' Evaluations of Educational Quality (SEEQ) questionnaire in Greek higher education</td>
</tr>
<tr>
<td><strong>Article Type:</strong></td>
<td>Original Research</td>
</tr>
<tr>
<td><strong>Keywords:</strong></td>
<td>educational evaluation; students' evaluations of teaching (SET); teaching effectiveness; SEEQ</td>
</tr>
</tbody>
</table>
| **Corresponding Author:** | Vasilis Grammatikopoulos, Ph.D.  
University of Crete  
Rethymnon, Crete GREECE |
| **Corresponding Author's Institution:** | University of Crete |
| **First Author:** | Vasilis Grammatikopoulos, Ph.D. |
| **Order of Authors:** | Vasilis Grammatikopoulos, Ph.D.  
Michalis Linardakis, Ph.D.  
Athanasios Gregoriadis, Ph.D.  
Vassilios Oikonomidis, Ph.D. |
| **Abstract:** | The aim of the current study was to provide a valid and reliable instrument for the evaluation of the teaching effectiveness in the Greek higher education system. Other objectives of the study were: (a) the examination of the dimensionality and the higher-order structure of the Greek version of Students' Evaluations of Educational Quality (SEEQ) questionnaire, and (b) the investigation of the effects of several background variables on Students' Evaluations of Teaching (SET) scores. 1264 students from eight universities participated by filling in the questionnaires administered to them. The results showed solid evidence of the applicability of the Greek version of SEEQ, by confirming the factor structure of the instrument and reassuring the multidimensionality of the teaching effectiveness construct. Additionally, the effects of several background variables on teaching effectiveness further supported the validity of SET scores. |
Assessing the Students’ Evaluations of Educational Quality (SEEQ) questionnaire in Greek higher education

Grammatikopoulos\textsuperscript{a}, V., Linardakis\textsuperscript{a}, M., Gregoriadis\textsuperscript{b}, A., & Oikonomidis\textsuperscript{a}, V.

\textsuperscript{a} University of Crete, School of Education, University campus, 7410 Rethymnon, Greece

\textsuperscript{b} Aristotle University of Thessaloniki, School of Education, School of Education building, 54124 Thessaloniki, Greece

Corresponding author:

Vasilis Grammatikopoulos, Ph.D.
University of Crete, School of Education
University Campus, 74100 Rethymnon, Crete, Greece
e-mail: gramvas@uoc.gr
tel: +30 28310 77655
fax: +30 28310 77654
Abstract

The aim of the current study was to provide a valid and reliable instrument for the evaluation of the teaching effectiveness in the Greek higher education system. Other objectives of the study were: (a) the examination of the dimensionality and the higher-order structure of the Greek version of Students’ Evaluations of Educational Quality (SEEQ) questionnaire, and (b) the investigation of the effects of several background variables on Students’ Evaluations of Teaching (SET) scores. 1264 students from eight universities participated by filling in the questionnaires administered to them. The results showed solid evidence of the applicability of the Greek version of SEEQ, by confirming the factor structure of the instrument and reassuring the multidimensionality of the teaching effectiveness construct. Additionally, the effects of several background variables on teaching effectiveness further supported the validity of SET scores.

Key words: educational evaluation, students’ evaluations of teaching (SET), teaching effectiveness, SEEQ
Assessing the Students’ Evaluations of Educational Quality (SEEQ) questionnaire in Greek higher education

Abstract

The aim of the current study was to provide a valid and reliable instrument for the evaluation of the teaching effectiveness in the Greek higher education system. Other objectives of the study were: (a) the examination of the dimensionality and the higher-order structure of the Greek version of Students’ Evaluation of Educational Quality (SEEQ) questionnaire, and (b) the investigation of the effects of several background variables on Students’ Evaluations of Teaching (SET) scores provided by the Greek version of SEEQ. 1264 students participated by filling in the questionnaires administered to them. The participants were selected from social science departments that belonged to eight universities of Greece. The results showed solid evidence of the applicability of the Greek version of SEEQ, by confirming the factor structure of the instrument and reassuring the multidimensionality of the teaching effectiveness construct. Additionally, the effects of several background variables on teaching effectiveness further supported the validity of SET scores.
Assessing the Students’ Evaluations of Educational Quality (SEEQ) questionnaire in Greek higher education

1. Introduction

1.1. The students’ evaluations of teaching effectiveness (SET) in higher education

Students’ evaluations of teaching (SET) are universal measures and have been applied in almost every higher educational system in the world (Zabaleta 2007). Universities have been developing and implementing SET instruments, analysing the results and using them for the evaluation of the provided instructional quality (Spooren, Brockx, and Mortelmans 2013).

The expansion of SET effectiveness in higher education the previous decades has triggered an on-going discussion about their validity. Aleamoni (1987, 1999) supported the validity of SET with his work ‘The Student Rating Myths versus Research Facts’ in which he presented common ‘myths’ about SET and claimed that faculties and administrators manufactured them in order to argue against the value of SET. On one hand, recent studies challenge the validity of SET (Feely 2002; Germain and Scandura 2005; Safer, Farmer, Segalla, and Elhoubi 2005). On the other hand, a plethora of well-designed studies confirmed the usefulness and validity of SET (Alsmadi 2005; Balam and Shannon 2010; Coffey and Gibbs 2001; Marsh 2007a, 2007b; Obenchain, Abernathy, and Wiest 2001).

There are also some recent review studies that did not offer support to either of the two opposite opinions (Jones, Gaffney-Rhys, & Jones 2014; Spooren et al. 2013). For example, in their extended literature review on the validity of SET, Spooren et al. (2013) could not draw firm conclusions. The researchers argued that SET validity are possibly affected by the variety of methods, measures and populations used in the studies, and pointed that in most cases, institutional designed instruments were used instead of widely accepted standardized...
instruments (e.g. Student Evaluation of Educational Quality, Course Experience Questionnaire). Another recent study investigated the effect of students’ personality traits on SET scores, finding a clear and consistent relationship between them (McCann and Gardner 2013), whereas Boysen, Kelly, Raesly, and Casner (2013) revealed the possible misinterpretation of SET in terms of generalization from limited data. Spooren et al. (2013) presented a very detailed overview of possible student, teacher and course—related characteristics that might affect SET scores. They indicated that not all characteristics could be considered as biasing factors, since some of them (e.g. student effort, class attendance, prior subject interest) are indicators of learning and are reasonably related to SET scores.

In addition to the existing arguments questioning the validity of SET, many researchers (e.g. Marsh, 1981, 1987, 2007a; Spooren et al. 2013; El Hassan 2009) argued that in most cases students actually evaluate the instructor of the course and not the instructional approach or the quality of learning. Thus, SET data have to be handled with caution and faculties also have to rely on other sources of information (d’Apollonia and Abrami, 1997; Marsh 2007a), especially if SET scores are going to be used for employment decisions (Jones et al. 2014). This contradictory context is furthermore empowered by research evidence supporting that there is not a single criterion for teaching effectiveness (Abrami, d’Apollonia, and Cohen 1990; Abrami and Mizener 1983; Marsh 1987, 1994, 1995; Marsh and Roche 1997), and thus the validity question is difficult to be universally answered.

The criticism on the validity of SET scores led to the inquiry of alternative sources for the evaluation of teaching effectiveness. Such sources might be committees of experts, peer-evaluation, self-evaluation, students’ grades etc. Yet, research evidence revealed that SET scores were more valid and reliable than any other source, and thus data from other sources shall be used together with SET scores and not instead of them (Cashin 1989, 1995; Chism 1999; Marsh 1987; McKeatchie 1997; Zhao and Gallant 2012).
The development of new instruments by each University might not be an ideal solution, because instruments that are applied only in one setting are not easily 'evaluated in relation to rigorous psychometric considerations and revised accordingly' (Marsh 2001, 4). The use of standardized and widely implemented instruments can result to valid and reliable SET scores. A lot of SET instruments have been developed and applied over the past few decades (see review of Spooren et al. 2013).

The instrument used in the current study is the Students’ Evaluations of Educational Quality (SEEQ). It is a questionnaire developed about 30 years ago (Marsh 1982) and it is considered one of the most widely used and universally accepted instruments. SEEQ has successfully provided valid and reliable SET scores in a variety of environments, such as several different higher education settings and different countries (e.g. Australia, USA, UK, Hong Kong, China, Spain, India) (Balam and Shannon 2010; Coffey and Gibbs 2001; Marsh 1986; Marsh, Hagu, Chung, and Siu 1997; Watkins and Thomas 1991).

Apart from its international recognition, another critical point for choosing SEEQ as the instrument of this study was the theoretical basis on which it was developed. Surprisingly, a lot of other existing SET instruments did not take under consideration the theories of teaching and learning in higher and adult education. Marsh and Dunkin (1992) evaluated the content of SEEQ in relation to general principles of teaching and learning in postsecondary education reported by Feldman (1976), Fincher (1985), and Mackie (1981). They revealed that SEEQ factors adequately included the principles described on the aforementioned studies. Moreover, the 'superiority' of SEEQ against other SET instruments also relies on psychometric analyses, as it constantly reveals high levels of validation and reliability scores (Coffey and Gibbs 2001; Marsh 1987; Marsh and Hocevar 1991a).

1.2. Evaluation of teaching effectiveness in the Greek higher educational setting
Until 2008, the Greek higher education had not established any official evaluation procedures for teaching effectiveness. Although academic staff, in order to be promoted in a higher rank or get appointed has to be evaluated for teaching effectiveness, these procedures were not officially established. A possible explanation is that in Greece, all academic promotions are open to any candidate, something that hampers the assessment of the teaching effectiveness of academics applying for the same position and coming from different universities, different countries or even different departments.

In 2005, the Hellenic Quality Assurance and Accreditation Agency for Higher Education (http://www.adip.gr) was founded, and one of the duties assigned to the Agency was the evaluation of teaching effectiveness in the Greek higher education.

The Agency adopted an internal evaluation procedure in 2007, and SET questionnaires were developed in order to assess teaching effectiveness. Yet, instead of choosing a standardized instrument, the Agency proposed some general factors and items and then let the university departments to freely adjust these instruments by adding, changing or removing items. This variety of SET instruments limited the possibility of comparing results and drawing overall conclusions. Marsh (2001) revealed that 'home-made' instruments are unlikely to evaluate the dimensions of teaching effectiveness broadly, and their usefulness can be questioned.

1.3. Multidimensionality and higher-order structure of SET

SEEQ is an instrument based on the perception that teaching effectiveness is a multidimensional construct and that SET scores have to be interpreted as such. This perception has been demonstrated in many studies conducted by Marsh (1982, 1984, 1987, 1991a, 1991b) and colleagues (Marsh and Dunkin 1992; Marsh and Hoceva, 1991a). However, there are also studies arguing against the multidimensionality interpretation of SET scores (Abrami and d'Apollonia 1991) or even in favor of both aspects. Apodaca and Grad
(2005) claimed that SET scores could be treated as multidimensional as much as unidimensional. This debate flourished because SET are used not only for feedback, but also for administrative purposes, where a single score representing an overall assessment of the instructional competence is very desirable.

Marsh (1991b) supported the higher-order structure of the teaching effectiveness notion, but also indicated that a unidimensional interpretation of SET scores could be possible, only by weighting the multiple dimensions of teaching effectiveness differently. Recent studies also provided support for the higher-order structure of teaching effectiveness (Apodaca and Grad 2005; Burdsal and Harrison 2008; Cheung 2000; Harrison, Douglas, and Burdsal 2004; Mortelmans and Spooren 2009). In their review study, Spooren et al. (2013) argued that SET instruments should cover the multidimensionality of the notion. They added that scores should be multidimensionally interpreted for feedback and formative purposes, but when it comes to unidimensionally interpretation they should better be weighed.

1.4. The effect of background variables on SET

A very frequently asked question has to do with the possible effect that several background variables might have on SET scores, the 'bias question' (Spooren et al. 2013). There are variables that cannot be considered as 'bias factors' as they are meaningful criteria of teaching effectiveness (e.g. class attendance or student effort), and their relation to SET scores implies valid teaching effectiveness effects rather than bias (Marshand Roche 1997, 2000; Spooren et al. 2013). Research interest is focused on whether factors that are not directly related to teaching effectiveness (e.g. instructor’s gender or course discipline) affect SET scores (Centra and Gaubatz 2000).

The discussion among researchers focuses mainly on whether some factors can be considered as bias (leniency hypothesis) or as meaningful dimensions (validity hypothesis) (e.g. the Expected Grade, or Workload/Difficulty factors) (Spooren et al. 2013). Leniency
hypothesis claims that instructors can 'buy' favourable evaluations by giving higher grades or reducing the workload and difficulty of a course (see e.g. Langbein 2008; McPherson and Todd Jewell 2007). On the other hand, many studies argued for the validity hypothesis, proposing that positive relationships, for example between SET scores and Expected Grade are generated by the quality of learning that occurred during the course, and therefore higher grades should be expected (Marsh and Roche 2000). Moreover, Marsh and Roche (2000) did not find any negative relationships between the Workload/ Difficulty of a course and SET scores. The current study adopted the validity hypothesis interpreting the data, as research has additionally revealed that 'bias' factors had little and not consistent influence on SET scores (Addison, Best, and Worrington 2006; Centra 2003; Marsh and Roche 2000).

Marsh (1987) and Marsh and Roche (1997) described four background variables as the most influential: (a) Prior Subject Interest, (b) Expected Grade, (c) Workload/ Difficulty, and (d) students selecting the course for General Interest. The design of the current study included the first three and the variable 'Professor rank.' The 'Professor Rank' was selected because some earlier studies provided contradictory results about it (Feldman 1983; Marsh 2007a; Marsh and Hocevar 1991b). The exclusion of the 'General Interest' factor was decided because the majority of the courses evaluated were not a subject of free choice of the students, due to the structure of their program study.

1.5. Purpose of the study

The main goal of the current study was to provide a valid and reliable instrument for the evaluation of teaching effectiveness in the Greek higher education. Additional objectives were: (a) to examine the dimensionality and the higher-order structure of the Greek version of SEEQ, and (b) to investigate the effects of several background variables on SET scores provided by the Greek version of SEEQ.

2. Method
2.1. Participants

1264 students from 41 different courses participated in the current study by filling in the questionnaires (SEEQ) administered to them, between the 10th and the 13th week of the spring semester of 2013. The participants were selected from social science departments from eight different Greek universities.

2.2. Instruments

The instrument used was the *Students’ Evaluations of Educational Quality* (SEEQ) questionnaire (Marsh, 1982, 1987, 1991b). The SEEQ comprises 35 items for measuring nine dimensions (Learning/Value-five items, Instructor Enthusiasm-five items, Organization-four items, Group Interaction-four items, Individual Rapport-four items, Breadth of Coverage-four items, Examinations/Grading-three items, Assignments/Readings-two items, and Workload/Difficulty-four items).

Prior to the study, SEEQ was translated into the Greek language by the authors. Then a bilingual academic, with great experience in the field translated the instrument into English again. The bilingual academic and the authors compared the two versions of the instrument (the original and the back-translated version), and any discrepancies found were corrected in the Greek version of the instrument. Next, the translated SEEQ was administered to some students and colleagues as a pilot test of face validity. Some changes were introduced in order to improve the meaning of some items in Greek.

2.3. Procedure

The period for the administration of the questionnaires was prearranged and while students were filling in the questionnaires, the instructor was always outside the classroom. Afterwards, a student collected the questionnaires, and put them in an envelope. The sealed envelope was then delivered by post to the authors. The students were informed that the
 instructor would not receive feedback about the evaluation of his/her teaching before the end of the semester and that this feedback would be only summarized comments.

2.4. Models tested

A Confirmatory Factor Analysis model (CFA) was conducted. Moreover, three higher order CFA models were postulated and tested based on the study of Marsh (1991b), named SEEQ Model 1 with one (general effectiveness), SEEQ Model 2 with two (skill, & rapport) and SEEQ Model 3 with three (presenter, rapport, & regulator) higher-order factors. These higher-order CFA models are all nested in the initial factor model, having each first-order factor load exclusively on a single higher-order factor.

2.5. Path analysis for the effect of background variables

The effect of a set of background variables on the SET scores of the Greek version of SEEQ was examined with path analysis. The set of the background variables that were included in the path model were Prior Student Interest in the Subject, Expected Grade, Professor Rank, and Workload/Difficulty. In the path model, the nine factors of the Greek version of SEEQ were included as latent variables.

2.6. Statistical analysis

CFA using Mplus v. 6.0 (Muthén and Muthén 1998-2010) was employed to confirm the factor structure, and to examine the viability of three postulated higher-order models of the Greek version of the SEEQ. A path analysis was used to investigate the effects of several background variables on SET scores provided by the Greek version of SEEQ. In all models, the maximum likelihood estimator with robust standard errors was used, due to non-normality of several items.

To evaluate the factor structure of the Greek version of SEEQ, the following fit indices were used: (a) the root mean square error of approximation (RMSEA) which is relatively insensitive to sample size and has better index to test the model fit compared with $X^2$ (Zhao
and Gallant 2012), (b) the standardized root mean square residual (SRMR), and (c) the comparative fit index (CFI). The cutoff values for the goodness of fit for the above indices were based on Hu and Bentler (1999) suggestions. Although they acknowledged that it is difficult to define specific cutoff value for any index, they indicated that values for CFI around .95, for SRMR ≤ .08, and for RMSEA ≤ .06 seem to be appropriate for determining goodness of fit.

To compare the three higher order models with the initial CFA model, besides the aforementioned indices (RMSEA, SRMR, & CFI), the Sattora-Bentler Scaled $\chi^2$ (Sattora and Bentler 2001) and the sample size adjusted BIC were used.

The effects of the background variables (Prior Student Interest in the Subject, Expected Grade, Professor Rank, and Workload/Difficulty) on the factors of the Greek version of the SEEQ were tested using path analysis.

3. Results

3.1. Confirmatory Factor Analysis of the Greek version of SEEQ

CFA results for the Greek version of SEEQ revealed acceptable values of the indices, supporting the proposed initial nine-factor structure of the original scale (Table 1).

Although the indices revealed reasonable fit to the data (Initial model of the Greek SEEQ, Table 1), it was decided to try to further improve the fit by adding the terms that the modification indices suggested (Final model of the Greek SEEQ, Table 1). Indeed, the model was further improved as it is revealed by the improved values of all fit indices presented in the Table 1.

Insert Table 1 about here

3.2. Higher-order structure of the Greek version of SEEQ

CFA results for the viability of three postulated models are presented in Table 2. The
structure of the three higher-order models is based on the structure of the first-order model, without the additional parameters that the modifications indices proposed, which also set the upper limit for the fit of the subsequent higher-order models (Marsh 1991b). Each of the higher-order models was compared to the first-order base model using the Sattora-Bentler test (Table 2).

Insert Table 2 about here

All three higher-order models revealed acceptable values for the indices used. Model 3 (three higher-order factors) appeared to be better than the remaining models positing one or two higher-order factors. Yet, the Sattora-Bentler Scaled $X^2$, that was used to compare the initial CFA model with the three higher order models, showed that in all three cases the null hypothesis was rejected, supporting the better fit of the first-order CFA model over the competitive models.

3.3. **Background variable effects on SET**

Path analysis was adopted to investigate the role of background variables on SET scores. The set of background variables comprised the variables Prior Student Interest in the Subject, Expected Grade, Professor Rank, and Workload/Difficulty, whereas the nine factors of the Greek version of SEEQ served as the latent factors of the path model. The results supported the fit of the model ($X^2 = 1663.232$, $df = 567$, $CFI = .950$, $RMSEA = .04$, and $SRMR = .039$).

Expected Grade affected positively all the nine factors of the Greek version of SEEQ. Professor Rank affected negatively the factors Instructor Enthusiasm, Group Interaction, and Individual Rapport, positively the factors Learning/Value, Organization, and Assignments/Readings, whereas no significant effects were detected on the factors Breadth of Coverage, Examinations/Grading, and Workload/Difficulty. Prior Student Interest in the Subject affected positively all the factors of the Greek version of SEEQ, whereas no
significant effect was detected on the Workload/Difficulty. Finally, Workload/Difficulty affected the factor Learning/Value negatively, the factor Examinations/Grading positively, whereas no significant effect was detected for the other factors of the Greek version of SEEQ. The results of the path analysis are presented in Figure 1.

4. Discussion

4.1. Psychometric properties of the Greek version of SEEQ

The basic aim of this study was to provide to the Greek higher education setting a valid and reliable instrument for the evaluation of teaching effectiveness. To this respect, the factorial validity of the Greek version of SEEQ was supported. The 35 items of the questionnaire comprised a nine-factor model exactly the same as the original SEEQ (Marsh 1982). The results added additional support to the strong psychometric properties of SEEQ across different cultures and environments (Balam and Shannon 2010; Coffey and Gibbs 2001; Marsh 1986; Marsh et al. 1997; Watkins and Thomas 1991). The findings are in accordance with the widely supported notion that SET instruments should capture multiple dimensions of teaching effectiveness (Spooren et al. 2013).

The provision of a valid teaching effectiveness instrument will help Greek researchers to overcome the limited value and usefulness of several 'home made' instruments used currently in Greece. The Greek version of SEEQ is an easy to use and widely accepted measure, and can serve as a basis for comparable findings in the Greek higher education settings.

4.2. Multidimensionality and higher-order structure

While other methods, such as peer-evaluation, committee of experts, or student learning outcomes are available to evaluate teaching effectiveness, SET still dominate the higher
education settings (Balam and Shannon 2010). In his seminal study, Marsh (1982) argued that students' evaluations of teaching effectiveness are best understood by multiple dimensions, as teaching shall be regarded a multidimensional construct. The results of this study confirmed the multidimensionality of the teaching effectiveness construct evaluated by the Greek version of SEEQ, as it was shown in other studies (Marsh 1982; Marsh 1991b).

The wide variety of dimensions captured in different SET instruments initiated the discussion about the existence of a single higher-order factor representing the teaching effectiveness contrast (Spooren et al. 2013). Additional pressure towards this direction was added by the use of SET scores for summative and administration purposes, where a single score capturing a general instructional skill would be very meaningful (Apodaca and Grad 2005). Abrami (1985) argued for a single higher-order factor model, whereas other research studies supported a multiple higher-order factors structure of teaching effectiveness (Apodaca and Grad 2005; Burdsal and Harrison 2008; Cheung 2000; Harrison, Douglas, and Burdsal 2004; Marsh 1991b; Mortelmans and Spooren 2009). The current findings provide further evidence against the single higher-order structure of the teaching effectiveness construct, revealing that a three higher-order factors model to be most tenable.

The Greek version of SEEQ higher-order structure differentiated slightly from the structure reported in Marsh's study (1991b), as a four higher-order factors structure was not supported. In that study (March 1991b), the three higher-order factor model was also a tenable model, yet slightly less sound than a four higher-order factor model. Additional research is expected to shed light in the number of the higher-order factors of the Greek version of SEEQ. The results of the current study supported the notion that even if a single score for summative purposes is more useful, SET scores are still multidimensional. A compromise is yet possible if weighted scores are going to be used (Marsh 1991b).

4.3. The effects of background variables
As it was mentioned before, the current study adopted a construct validity approach. The leniency hypothesis was not supported by the results, as the Workload/Difficulty dimension has no significantly effect on six out of eight dimensions of the Greek version of SEEQ. Workload/Difficulty significantly affected only the factors Learning/Value (-.067, p < .05) and Examinations/Grading (.071, p < .05), but these low values could not support the leniency hypothesis. The negative correlation between Workload/Difficulty and Learning/Value might be attributed to the fact that students with excessive workload are unlikely to absorb the provided material and as a result, learning 'suffers' (Marsh and Roche 2000).

The Expected Grade’s significant effects on the variables of the Greek version of SEEQ also supported the validity hypothesis. The highest effect was on Learning/Value variable (.404, p < .01) indicating that better expected grades might reflect better learning by students as Expected Grade actually reflects a measure of learning (Marsh and Roche 1997). Additional support was revealed by the negative correlation between Expected Grade and Workload/Difficulty (-.293, p < .01). This is an a-priori reasonable prediction, as students’ expectations for grades are unlikely to be high due to the difficulty and heavy load of a specific course.

In our study, Prior Subject Interest mostly influenced Learning/Value as in Marsh and Roche (1997) study. Moreover, its positive effect on eight dimensions of the Greek version of SEEQ supported the validity hypothesis, as students with higher prior interest in the course shall be expected to provide more positive SET ratings (Marsh and Roche 1997). The only factor that was not significantly correlated with Prior Subject Interest was the Workload/Difficulty factor. This is also an a-priori reasonable prediction, as the interest in an academic subject and the difficulty of a course are unlikely to be correlated.

The fourth background variable tested was Professor Rank. It can be assumed that
younger instructors (lower rank) would be more enthusiastic, provide better interactions and create a more positive climate than the older ones (higher rank) (Feldman 1983). Indeed, the current study’s findings revealed negative correlations between Professor Rank and Enthusiasm, Group Interaction, & Individual Rapport. On the other hand, the positive correlations with Learning/Value, Organization, & Assignments/Readings can be attributed to the accumulated experience of higher ranked Professors. These results are in contradiction with Marsh and Hocevar (1991b), and Marsh (2007a) studies, where added experience did not influence teachers’ effectiveness. To this respect, further research efforts are needed to explore the relation between the staff’s experience and SET effectiveness in the Greek educational settings.

In conclusion, the current study has theoretical and practical importance. This study provided solid evidence for the applicability of SEEQ in the Greek higher education, by confirming the factor structure of the instrument. It also provides the Greek higher education setting with a valid instrument that can be used in a variety of environments for internal evaluation purposes. Moreover, the results reassured the sound psychometric properties of SEEQ indicating the value and usefulness of the instrument.

4.4. Limitations of the study

The structure of the program of study in the Greek universities, and the limited sources in terms of financial support were responsible for some limitations of the current study.

a) The role of the important background factor of students selecting a course for “General Interest” (Marsh 1987, Marsh and Roche 1997) could not be incorporated in the current study because of the compulsory nature of the most courses selected in the current study. Additionally, the sample of the current study relied only on courses at social sciences departments.
b) Several studies revealed the importance of multiple sources of information to better capture the SET construct. Yet, the findings of the current study relied only on the use of students’ evaluations of teaching effectiveness, as no other measure evaluating teaching effectiveness was used (e.g. teacher evaluations, peer-evaluation, panel of experts).

c) The underlying structure of the Greek version of SEEQ was based on the back translation and the CFA adopted in the current study, evaluating only the face and structural validity.

4.5. Conclusions

Evaluating teaching effectiveness in higher education is not an easy task due to the complexity of the construct. Even if the results of the current study were promising, additional research will help to better understand the SET construct in the Greek higher educational settings. Future efforts must select a wider variety of departments towards a representative sample, and investigate the role of more background factors of the teaching effectiveness construct. Additional resources evaluating teaching effectiveness will shed more light on the SET construct by providing comparable data. Finally, future research studies must further investigate others aspects of validity such as 'evaluating test validity is not a static, one-time event; it is a continuous process' (Mertens 2010, 384).

5. References


Figure 1. Path diagram: Effects of the background factors on the scores of the Greek version of SEEQ

* significance level at .05
** significance level at .01
Table 1. The fit of the Greek version of SEEQ

<table>
<thead>
<tr>
<th></th>
<th>( X^2 ) (df)</th>
<th>Sample sized adjusted BIC</th>
<th>CFI</th>
<th>RMSEA</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial model</td>
<td>1934.382 (491)</td>
<td>133948.583</td>
<td>.932</td>
<td>.048</td>
<td>.043</td>
</tr>
<tr>
<td>Greek SEEQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final model</td>
<td>1455.171 (486)</td>
<td>133302.781</td>
<td>.954</td>
<td>.040</td>
<td>.039</td>
</tr>
<tr>
<td>Greek SEEQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Model comparison between first-order CFA and SEEQ with one, two, and three higher order factors

<table>
<thead>
<tr>
<th></th>
<th>$\chi^2$ (df)</th>
<th>Scale Corr. Factor for MLR</th>
<th>Sample sized adjusted BIC</th>
<th>CFI</th>
<th>RMSEA</th>
<th>SRMR</th>
<th>Sattora-Bentler Scaled $\chi^2$ (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Initial model</strong></td>
<td>1934.382 (491)</td>
<td>1.365</td>
<td>133948.583</td>
<td>.932</td>
<td>.048</td>
<td>.043</td>
<td></td>
</tr>
<tr>
<td><strong>Model 1</strong></td>
<td>2339.750 (518)</td>
<td>1.372</td>
<td>134410.571</td>
<td>.914</td>
<td>.053</td>
<td>.056</td>
<td>379.982 (27)</td>
</tr>
<tr>
<td>(HO = 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model 2</strong></td>
<td>2285.658 (517)</td>
<td>1.371</td>
<td>134339.184</td>
<td>.916</td>
<td>.052</td>
<td>.055</td>
<td>332.280 (26)</td>
</tr>
<tr>
<td>(HO = 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model 3</strong></td>
<td>2189.774 (515)</td>
<td>1.370</td>
<td>134212.757</td>
<td>.921</td>
<td>.051</td>
<td>.054</td>
<td>244.217 (24)</td>
</tr>
<tr>
<td>(HO = 3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>