Promoting deep approach to learning and self-efficacy by changing the purpose of self-assessment: A comparison of summative and formative models

Juuso Henrik Nieminen[a]*, Henna Asikainen[b], Johanna Rämö[a]

[a] *Department of mathematics and statistics, University of Helsinki, Helsinki, Finland*

[b]*Department of biosciences, University of Helsinki, Helsinki, Finland*

*juuso.h.nieminen@helsinki.fi

orcid.org/0000-0003-3087-8933

Promoting deep approach to learning and self-efficacy by changing the purpose of self-assessment: A comparison of summative and formative models

Self-assessment has been portrayed as a way to promote lifelong learning in higher education. While most of the previous literature builds on the idea of self-assessment as a formative tool for learning, some scholars have suggested using it in a summative way. In the present study, we have empirically compared formative and summative models for self-assessment, based on different educational purposes (N = 299). Latent profile analysis was used to observe student subgroups in terms of deep and surface approaches to learning. The results show that the student profiles varied between the self-assessment models. The students taking part in the summative self-assessment group were overrepresented amongst the profile with high level of deep approach to learning. Also, summative self-assessment was related to an increased level of self-efficacy. The study implies that summative self-assessment can be used to foster students' studying; however, this requires a context where aligning self-assessment with future-driven pedagogical purposes is possible.

Keywords: self-assessment; summative assessment; formative assessment; approaches to learning; self-efficacy

Introduction

It is often stated that the fundamental goal of higher education (HE) is to prepare students for lifelong learning by taking responsibility for their own learning (Boud and Falchikov 2006). As Levine and Dean (2012) point out, we are educating university students in an era of continuing change, which underlines the importance of teaching deep learning methods in contrast to fragmented pieces of information. However, these fundamental goals of HE are not always seen in the assessment practices. Studies have shown that traditional assessment methods still dominate in HE (Beaumont, O'Doherty, and Shannon 2011; Postareff et al. 2012), and further, the current practices tend to overemphasise the importance of assessment for certification and validation purposes

(Crisp 2012). Currently, it can be argued that in general there is a gap between what is valued in HE and how students are assessed. Traditional assessment methods, such as individual exams, are known to not always support the goals of ‘lifelong learning’ (e.g. Knight 2002).

In the present study, student self-assessment (SSA) has been used to support the quality of studying and to express the educational goals of HE. The literature on SSA differentiates between *self-assessment* and *self-grading* (Andrade and Du 2007). Self-assessment refers to a formal process during which students make judgements about their own learning and compare it with explicitly-stated criteria (Panadero, Brown, and Strijbos 2016; Tan 2008). According to Andrade and Du (2007), self-grading is seen as a method that involves students in grading their own work. The present study connects the concepts of self-*grading* and self-*assessment* with the ones of summative and formative assessment. Summative assessment practices are used after the learning process to determine what the students know and to ensure student comparability (Shute and Kim 2014). Formative assessment, on the contrary, refers to assessment that seeks to improve and accelerate students’ learning through continuous feedback (Broadbent, Panadero, and Boud 2018). However, both formative self-assessment and summative self-grading practices should not only be seen as practical methods; their underlying pedagogical purposes should be concerned as well.

In the present study, two different ways of conducting SSA (self-assessment *models*), based on formative and summative purposes, were compared drawing on person-oriented analysis, bringing research-based evidence to the field. The purpose was to examine whether there are differences between formative (self-*assessment*) and summative (involving self-*grading*) models of SSA in terms of how students study;

approaches to learning, self-efficacy beliefs and course achievement were used as indicators. Next, the theoretical basis for formative and summative SSA are introduced.

Self-assessment as a formative tool for learning

Broadly, SSA has been defined as involving students' own monitoring on their work or process (Brown and Harris 2013). In the previous literature, self-assessment has mainly been recommended for use as a formative tool for learning (Andrade and Cizek 2010; Andrade and Du 2007; Brown and Harris 2013; Panadero et al. 2016); this fits with the previously introduced definition of *self-assessment*. This kind of SSA means that the students reflect on their own learning based on pre-set learning criteria during the learning process. In educational settings, formal self-assessment tasks can be based either on *rubrics* (that communicate the learning objectives in a form of a matrix) or on *scripts* (a set of questions asking the students to reflect on their learning) (Alonso-Tapia and Panadero 2010; Panadero, Tapia, and Huertas 2012).

The idea in formative SSA is that students benefit from it, even though teacher is responsible for the last word - the grade (Bourke 2018). Through formative SSA, with feedback provided, students learn to calibrate their own ideas about their skills with the learning objectives (Panadero et al. 2016). Formative self-assessment has also been reported as promoting learning of a higher quality (Andrade and Du 2007; Brown and Harris 2013; Panadero et al. 2012) and improved motivational factors (Andrade and Du 2007). The previous literature also supports the view that formative self-assessment practices possess an opportunity to enhance learning and should, therefore, be used in addition to more traditional assessment practices.

Why is self-assessment only recommended for use in a formative way? These claims are not always based on empirical data. It has been suggested that 'human nature' (Andrade and Cizek 2010; Andrade and Valtcheva 2009) will make students dishonest and, therefore, only formative use of SSA is recommended. Concern about the validity of self-grading is often reported in the literature (e.g. Brown et al. 2015). Rarely has research provided such clear implications on practice: ‘Do not turn self-assessment into self-evaluation by counting it toward a grade’ (Andrade and Valtcheva 2009, 17). A similar view is shared by Bourke (2018), who claims that self-grading results in focusing on the grades, not on the learning; however, no data or scientific references have been offered to support this statement. However, an empirical study found that when students had a chance to evaluate 5% of their final course grade, their accuracy in self-assessment decreased (Tejeiro et al., 2012). Based on this, Tejeiro and colleagues suggest that SSA should only be used in a formative way. They identified cheating and emotional stress as barriers to honest and reflective self-assessment process.

As some studies suggest that HE students are not always competent to assess their own learning (e.g. Tejeiro et al. 2012), it is necessary to let the students practise their self-assessment skills (Panadero et al. 2016) and to offer them feedback on these skills. Usually SSA is used as part of a larger feedback cycle (Beaumont et al. 2011) in which self-assessment is only one of several feedback methods. The idea is that the students gain information about their learning through formative assessment and feedback. To conclude, it can be said that formative SSA ensures that students are involved in every step of assessment (Tan 2007).

Self-assessment as a summative, future-driven act

Contrary to the suggestions about using self-assessment only in a formative way (e.g. Bourke 2018), some scholars have suggested that effective SSA programs not only allow students to compare their work against a set of criteria but also to give them power over assigning their own grade (Strong, Davis, and Hawks 2004; Taras 2015, 2008). This idea relates to Andrade and Du’s (2007) definition of self-grading.

However, understanding self-assessment as a summative act does not simply mean self-grading at the end of a learning process but requires reconceptualisation of the whole purpose of assessment. Self-grading does not have to mean that students alone should grade themselves, but rather that the students have the ultimate power to reflect on the external feedback they receive from their teachers and peers. Summative SSA builds on formative SSA and therefore on feedback cycles (Beaumont et al. 2011). Self-grading, done only after students have actively engaged with formative SSA tasks, is seen as a 'process within a process in which many thoughtful and fair decisions have to be made according to pre-established and reasonably set criteria' (López-Pastor et al. 2012, 454).

We see summative SSA as being closely tied with the concept of future-driven self-assessment (Tan 2009, 2007). That means self-assessment aimed at developing the skills of lifelong learning; namely, skills that could be used outside the classroom. Tan (2009, 2007) sees future-driven SSA as a framework that calls for active learner agency. According to him, this can be accomplished by teaching students not only to compare their own self-assessed marks with the marks graded by teacher, but also by teaching them to evaluate their own judging abilities critically. Assessment methods that see students as active agents in the learning process are also emphasised by Boud and Falchikov (2006), who state that this is crucial to sustainable learning since 'neither teachers nor curriculum drives learning after graduation' (402). Summative SSA builds on these views, as the feedback provided by the teacher is only a base for reflection, while the students themselves have the power to evaluate whether they have reached the learning objectives for the grade they claim (Taras 2015, 2008). Thus, the objective of summative SSA is to teach evaluation skills for the future where there are no teachers or programmes to tell whether learning has happened. This is not to say that active learner agency wouldn't be a part of formative self-assessment as well. However, summative

SSA asks the students to take responsibility by giving them power over their grade, which might lead to a different kind of student agency. Whether summative and formative SSA affect studying in different ways falls exactly within the scope of the present study.

We were able to identify few empirical studies in which students were given power over their self-assessment by letting it count towards their grade. Friess and Davis (2016) report a study in which students either self-graded their homework submissions or took a quiz. The self-grading students showed better time management and they also reported self-grading as a more effective learning method than taking the quiz. Strong and colleagues (2004) graded their students but let them decide on their final grade by themselves. What they reported was an increase in student motivation and in the responsibility that the students took for their own learning. Also, Tejeiro and colleagues (2012) let their students decide 5% of their final grade. In their study they concluded that self-grading lowered the accuracy of SSA and therefore shouldn't be used in a summative way. It can be concluded that even though Andrade and Du (2007) suggest that confusing self-assessment with self-grading is common, there have been few empirical articles about using self-grading in HE.

The interaction between self-assessment and studying

The present study empirically compares studying by students taking part in formative and summative SSA. Three indicators for studying were used: approaches to learning, self-efficacy beliefs and course achievement. In this section, these concepts and their importance to studying are explained, as well as their connection with self-assessment.

Approaches to learning

In the present study, the underlying assumption is that there are always student subgroups that differ in how they benefit from self-assessment as an assessment method. Here, students' approaches to learning tradition (Asikainen and Gijbels 2017; Entwistle 2009) are used as a theoretical background to observe these subgroups. Traditionally, approaches to learning have been divided into the deep approach to learning, which emphasises aiming to understand and applying critical thinking, and the surface approach to learning, which emphasises memorising and struggling with the fragmented knowledge base (Asikainen and Gijbels 2017; Entwistle and Ramsden 1983). Usually, the deep approach to learning has been shown to be related to better learning outcomes than the surface approach has (Diseth 2003; Entwistle and Ramsden 1983). However, the dichotomy of the surface and deep approaches to learning is not straightforward; students may also apply different combinations of approaches to learning (e.g. Parpala et al. 2010).

As approaches to learning are situational and only exist in relation to learning environments (Richardson, Abraham, and Bond 2012), there has been a voluminous amount of research concerning whether it is possible to promote deep approach to learning. Often, assessment is seen as the answer. It has even been suggested that assessment is the main factor influencing students' approaches to learning (Rust, O'Donovan, and Price 2005). Results on how alternative assessment methods, such as peer- and self-assessment, affect approaches to learning are varied. Alternative assessment methods have been seen as a way to discourage passive learning rather than as a way to support deep approach to learning (Baeten, Dochy, and Struyven 2008; Struyven, Dochy, Janssens, and Gielen 2006). Further, alternative assessment has been linked to increased use of the surface approach (Gijbels and Dochy 2006). Gijbels and Dochy underline that students' perceptions of assessment are the key element in understanding these kinds of results. For example, if workload is perceived as being too high, students might prefer to use surface-oriented study methods. It has been suggested that students adapting a deep approach to learning might prefer alternative assessment methods that support learning (Baeten et al. 2008; Gijbels and Dochy 2006) and that students using the surface approach to learning might have a hard time adapting to assessment methods that favour the deep approach (e.g. Marton and Säljö 1976).

Previous studies often concluded that supporting the deep approach to learning with assessment causes profound difficulties (e.g. Struyven et al. 2006). Haggis (2003) even raised the question of whether the deep approach to learning could even be ‘induced’ if it is not ‘already there’ (94). However, some guidelines have been given for assessment that supports deep learning. Struyven and colleagues (2006) highlight the importance of feedback and structural support during assessment. Sadler and Good (2006) found that alternative assessment was able to support deeper understanding of the subject matter in middle school when assessment was not introduced as an isolated practice but was aligned with the educational purposes of the classroom. To sum up, there appears to be a research gap in what kind of assessment (and especially self-assessment) could support deep approach to learning.

Self-efficacy beliefs

In addition to having a great impact on students’ learning processes, self-assessment can also influence students’ self-efficacy beliefs. Students’ self-efficacy beliefs can be defined as one’s beliefs about one’s abilities to achieve in a given form of attainment (Bandura 1997). Self-efficacy beliefs have a great influence on performance and learning. Bandura (1997) argued that students with strong self-efficacy beliefs set higher goals and put more effort into their studying. A systematic review and meta-analysis exploring psychological correlates on university students’ performance showed that of 50 correlates affecting student performance, self-efficacy was the strongest predictor of academic performance (Richardson, Abraham, and Bond 2012). In addition, previous studies have shown that self-efficacy beliefs are related to students approaches to learning. Stronger self-efficacy beliefs have been found to be related to the deep approach to learning and weaker self-efficacy belief to the surface approach to learning (Diseth 2011; Prat-sala and Redford 2010). Students who believe they can succeed are also more likely to apply deeper processes of understanding in their learning.

Studies have shown that self-assessment can have a great positive impact on self-efficacy beliefs (e.g. Panadero et al. 2017; Panadero and Romero 2014). Panadero and colleagues (2017) stated that the reason for this can be that by obtaining deeper insights of the requirements of the task, students are more likely to succeed and experience successful performance. According to Bandura (1997), self-efficacy beliefs can be developed through experiences of mastering or being successful in a task. Thus, experiences of successful performances in self-assessment can also promote students' self-efficacy beliefs. Although the relationship between self-assessment and self-efficacy beliefs has been studied before, there is a gap in exploring self-efficacy beliefs in relation to different self-assessment practices.

Achievement on the course

Some earlier studies (e.g. Ibabe and Jauregizar 2010; Jay and Owens 2016) have suggested that self-assessment relates to higher learning results through students' active engagement in their own learning process. Therefore, we measured academic achievement in our study to see whether performance varied between the summative and formative self-assessment models.

Objectives of the study

The objective of the study was to examine empirically how students' studying (indicated by approaches to learning, self-efficacy and mathematical achievement) differ in two self-assessment models: formative and summative. The study used a person-oriented approach to explore student subgroups regarding deep and surface approaches to learning. The research questions were stated as follows: (1) What differences in approaches to learning, self-efficacy and mathematical achievement are there between the two self-assessment models? (2) Which student subgroups can be found from the whole student population in terms of approaches to learning? How are these subgroups represented in each of the self-assessment groups? (3) In each of the student subgroups, what differences are there regarding approaches to learning, self-efficacy and mathematical achievement in the two self-assessment models?

Context and the study design

The present study was conducted as a part of the Digital Self-Assessment (DISA) project at the University of Helsinki. An undergraduate mathematics course in a research-intensive university in Finland was designed for the study (see Figure 1). The five credit course (European Credit Transfer and Accumulation System) lasted for seven weeks. There were 426 participants at the beginning of the course, of which 313 were actively engaged and completed the final assessment. The topic of the course was linear algebra; it is one of the first courses mathematics students take. Overall, the course was designed to be student-centred. Teaching was based on the Extreme Apprenticeship Model (Rämö, Reinholz, Häsä and Lahdenperä 2019). It is a teaching model in which students take part in activities resembling those of experts. The Moodle online learning environment was used during the course.

The course was graded on a scale from 0 (‘fail’) to 5. It should be noted that in Finnish universities, grades do not determine students’ educational paths. Exams can usually be retaken multiple times, and grades are rarely asked for by future employers. Also, the Finnish Universities Act (2009) provides academic freedom for teaching and assessment methods.

At the beginning of the course, the participants were randomly divided into two groups: half of the students attended a course exam at the end of the course (*formative SSA group*, studying with the formative self-assessment model), while the other half self-graded themselves (*summative SSA group*, studying with the summative self-assessment model). Both groups took part in the same SSA practices during the course. Also, both groups were motivated to self-assess by telling them that learning how to evaluate one’s own work is an important skill and that the students should use the opportunity to learn for themselves, not for the teacher. Only the final summative assessment method was different for the two groups; otherwise, both groups experienced the same learning environment. Finally, after the final summative assessment, the data collection was conducted with a survey. Next, how the two self-assessment models were implemented in the practice is explained (Figure 1).

The formative self-assessment model in practice

The students in the formative SSA group (N = 147) took part of SSA tasks during the course; however, these self-assessments did not count towards their grade. The final summative assessment was conducted with a course exam. To support students’ self-assessment, the course utilised a detailed rubric to communicate the learning objectives. Some topics in the rubric were content-specific, such as ‘solving linear systems’, while others concerned generic skills, such as ‘reading and writing mathematics’. Examples of the learning objectives are given in Table 1. Of the topics, five concerned mathematical content and four concerned generic skills. The criteria were given for grades 1–2, 3–4 and 5.

The students completed two compulsory self-assessment tasks during the course. In the first task, the students were shown all the learning objectives that they had worked on so far. For each objective, they stated whether they felt they mastered it (1) well, (2) partially or (3) not yet. Also, by using scripts (Panadero et al. 2012), the students were asked to reflect in writing how they were doing and what their goals were. In the second SSA task, the students had to decide what grade they would award themselves from each topic in the rubric. Again, questions were asked about the students' feelings and goals. Also, the students had a chance to justify in writing their self-assessment for each of the learning objectives.

The course largely utilised feedback cycles (Beaumont et al. 2011) to support students' formative self-assessment. Digital feedback on students' self-assessments was offered. Each of the tasks in the course was linked with the learning objectives it was supporting, and based on the number of the tasks completed, the students received a computed index that indicated how well their self-assessment was in line with the work they had done during the course. It was explained to them that the indices were not necessarily representative of their skills, and they were encouraged to explain in writing if they believed that the coursework would not adequately reflect their skills.

Feedback cycles were also used with the mathematical tasks during the course. New topics were introduced through scaffolded tasks. Each week, students were given three sets of mathematics tasks, each representing a different kind of feedback. First, there were digital tasks offering automatic constructive feedback. Also, there were pen-and-paper tasks, which were divided into two sections. The first section comprised two or three tasks concerning the most central topics of the course. One of these tasks was selected for feedback that was provided by the student tutors who had been taught to write constructive feedback. Students had an opportunity to return a revised solution twice. The second section of pen-and-paper assignments consisted of tasks for which no feedback was provided; model answers for these tasks were published later.

During the course, students were offered guidance in an open drop-in learning space by student tutors who were trained for effective teaching methods. The learning space offered an opportunity for social interaction and for peer feedback. Also, digital peer assessment on mathematical tasks was provided on Moodle, and digital feedback on students' peer assessments was offered according to how constructive they were.

The summative self-assessment model in practice

The students in the summative SSA group (N = 152) took part in the same learning environment as the students in the formative SSA group. The only difference was the final summative assessment method. Therefore, the previous description of the feedback cycles concerns this group as well.

While the formative SSA group took part in the course exam, the students in the summative SSA group took part in the self-grading process. At the end of the course, students in the summative SSA group self-graded themselves in the same manner as in the second SSA task: grading was based the topics in the rubric. For each grade, students could reflect on why they chose that grade, in writing. They also awarded themselves the final grade. No instructions were provided on how the summative SSA group should arrive at the final grade.

The digital feedback system, normally used to offer feedback on students' self-assessment, was utilised at the end of the course to check the self-graded marks before their final validation. This was done to ensure that students with low self-efficacy would not assess themselves with a very low grade and to prevent obvious cheating. At the beginning of the course, all the students were told that the validation system was used only to prevent obvious cheating and not to reduce their power over their own grades. The system pointed out the students whose self-assessed and computed grades differed by more than one grade. There were 32 such students, and their grades were dealt with separately by the teacher responsible for the course. Of these students, 14 assessed themselves as very high in relation to their achievement during the course; the other 18 were either able to keep their self-graded mark or raise it if it was much lower than what the system implied.

Methodology

Instruments

Students' approaches to learning were measured with the HowULearn questionnaire (Parpala and Lindblom-Ylänne 2012) which has been shown to be a reliable measure in the context of Finnish HE (e.g. Herrmann, Bager-Elsborg, and Parpala 2017). We used two scales from the students' approaches to learning section: deep approach to learning (four items); and surface approach to learning (four items). Furthermore, self-efficacy was measured with the five-item scale from the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich 1991).

Students' score for achievement in the course was based on the scores of the three mathematical task sets: (1) tasks with automatic feedback (2) tasks with feedback from the student tutors (3) tasks with no feedback. The following formula was used:

$$\text{Achievement} = \frac{\text{total(set 1) / max(set 1) + total(set 2) / max(set 2) + total(set 3) / max(set 3)}}{3}$$

It is important to note that the present study only used these teacher-generated tasks as the measurement for 'achievement'. The achievement score should therefore only be seen as indicative for learning and studying during the course.

Participants

All the 313 students who completed the final assessment of the course were asked to take part in the study. A total of 302 students completed the survey after the course and gave their permission for us to use both their survey and course data in the research, with the response rate of 96.5%. Three students were excluded from the data since they hadn't answered the questions in the HowULearn instrument, thus resulting to the final N of 299 students. There were 152 students in the summative SSA group and 147 students in the formative SSA group (Table 2).

Age (M_{age} = 24.37, SD = 7.02, median = 21) showed no differences ($t(291)$ = .084, p = .933) between the summative (M_{age} = 24.40, SD = 6.72, median = 22) and formative SSA groups (M_{age} = 24.33, SD = 7.35, median = 21). Also, no differences were found between the groups regarding major of the studies ($\chi^2(9, N = 299) = 5.18$, p = .82; 24 majors were represented, and 94 students majored in mathematics) or gender ($\chi^2(3, N = 299) = .35$, p = .95). The groups did not differ in terms of achievement either, measured by course tasks with feedback of various types: automatic feedback ($t(292)$ = –.80, p = .42), tutor-led feedback ($t(296)$ = .88, p = .38) and the tasks with no feedback ($t(296)$ = –.53, p = .60). Overall, it can be stated that the student population in the study was homogeneous, and no differences were found in terms of the categorical variables of the study.

Analysis methods

The analysis of the study was divided into four stages. First, confirmatory factor analysis was conducted on the scales measuring deep and surface approaches to learning to ensure the construct validity of the research instrument. The fit for the model was based on Comparative Fit Index (CFI) and Root Mean Square Error of

Approximation (RMSEA). A good fit was indexed with CFI values above .95 and RMSEA values below .06 (Hu and Bentler 1999). A general comparison of the two SSA groups was conducted using t-testing (RQ1).

Latent profile analysis (LPA) was conducted with Mplus 8.0 on the whole student population to map out student subgroups regarding approaches to learning (RQ2). LPA offers a person-oriented analysis to classify individuals into homogenous subgroups by latent, underlying classes (Collins and Lanza 2010). The number of the profiles is presumed to be unknown, and the membership of a profile is assumed to explain the scores of continuous scales. LPA offers fit indexes for different cluster solutions, unlike some other clustering methods like hierarchical cluster analysis. Six fit indexes were used to compare between different profile solutions: Akaike Information Criterion (AIC; Akaike 1987), Bayesian Information Criterion (BIC; Schwarz 1978), the BIC Sample-Size Adjusted (aBIC), the Vuong-Lo-Mendell-Rubin Likelihood Ratio Test and the Lo-Mendell-Rubin Adjusted Likelihood Ratio Test (LMR LRT; Lo, Mendell, and Rubin 2001). Also, the size of the smallest profile and the interpretability of the profile solution were considered in the analysis.

The distribution of the students' profiles was compared with a Chi square test between the two SSA models (RQ2). Finally, t-testing within the profiles was conducted regarding approaches to learning, self-efficacy and course achievement (RQ3). Throughout the analysis process, missing values were treated as nulls.

Results

A general-level comparison of the self-assessment groups

The confirmatory analysis conducted on two scales measuring deep and surface approaches to learning had an acceptable fit (CFI = .96, RMSEA = .07). The indexes showed that one item measuring the surface approach (*'often I had to repeat things to learn them')* did not fit in the model. In addition, Spearman correlation analysis showed that all the other items measuring surface approach to learning correlated negatively with items measuring deep approach, but there was no relationship between this item. Thus, a second model with a good fit (CFI = .98, RMSEA = .04) was conducted with only three items in the surface approach. The reliability analysis showed that the consistency of the scale with three items (α = .75) did not differ much from the model with four items (α = .76). Thus, the three-item scale was chosen for this study. The reliability analysis measuring approaches to learning and self-efficacy scales showed a good level of consistency (α = 0.75 - 0.92). The homogeneity of variances of the variables were also tested: Levene's test indicated equal variances for all the variables (F = 1.17 … F = 2.25; $p > 0.05$) except for self-efficacy (F = 10.1, $p < .000$).

Descriptives of the variables in the two self-assessment models are shown in Table 3. A t-test analysis showed that the surface approach to learning was reported as being significantly more in the formative SSA group ($t(297) = –2.5$, $p = .013$, $d = .37$), while the deep approach to learning was reported as being significantly more in the summative SSA group ($t(297) = 3.26$, $p < 0.001$, $d = .29$). However, the effect sizes were only small or moderate. In addition, self-efficacy was reported to be significantly higher in the summative SSA group with a larger effect size ($t(297) = 5.03$, $p < 0.001$, $d = .59$).

Person-oriented view: Observing the student profiles

After conducting latent profile analysis in terms of deep and surface approaches to learning with the whole student population, various fit indexes were compared. Unsurprisingly, different indexes favoured different profile solutions (Table 4). While the AIC and aBIC indexes seemed to favour as small profiles as possible, the BIC index slightly favoured the solution with four profiles. The VLMR and LMR LRT indexes both favoured solutions with four (p_{VLMR}, $p_{LMR\ LRT} < .05$) and five (p_{VLMR}, $p_{LMR\ LRT} < .05$) profiles.

Finally, the results were also interpreted according to profile size. The solutions with five and six profiles included a very small student cluster (1 and 5 students, respectively). The solution with just two profiles was not selected since it would not truly differentiate between student groups. Finally, based on the fit indexes, interpretability and suitable-sized smallest profiles, the solution with four profiles was used in this study.

In the first profile, *students applying a very deep approach* (N = 116), students' scores on the deep approach were very high (Mean = 4.07; SD = .67) and their scores on the surface approach were really low (Mean = 1.28; SD = .26). This indicates that these students were predominantly studying in a way that reflects a will to have a deep understanding of the content rather than memorising it. In the second profile, *students applying a deep approach* (N = 116), students' deep approach scores (Mean = 3.57; SD = .69) were slightly higher than the average of the whole sample, and surface approach scores (Mean = 2.14; SD = .14) were likewise slightly lower than on average. What characterised the third profile, *students applying a dissonant approach* (N = 52), was that the students reported using both deep and surface approaches. These dissonant or incongruous profiles are often found in studies concerning approaches to learning (e.g. Lindblom-Ylänne 2003), making it an interesting profile to study. The smallest student cluster, *students applying a surface approach* (N = 15), consisted of students who reported high scores on surface approach to learning (Mean = 4.04; SD = .35); however, the scores on deep approach (Mean = 3.12; SD = .79) were only slightly lower than in the *dissonant approach* profile.

The profiles were characterised regarding self-efficacy and achievement in the course (Figure 2). Finally, ANOVA was conducted to observe differences in the study variables (Table 5). There were significant differences regarding all of the variables of the study, with effect sizes varying from medium (achievement: .14) to extremely large (surface approach: .89). Tukey's post hoc testing showed that students in the deep approach profile reported higher levels of self-efficacy than those in the other profiles and outperformed them in terms of achievement. Because the surface approach profile was small (N = 15) and since the variance of self-efficacy was unequal in the student profiles, nonparametric testing was also conducted. The Kruskal-Wallis test further validated the significant differences between the student profiles regarding all the study variables ($p < 0.001$).

The SSA models and the student profiles

The distribution of student profiles in the two self-assessment models is shown in Figure 3. A Chi-square test of independence was calculated comparing the profile formation in the two SSA models. A significant difference was found in the distributions of the profiles in the two models ($\chi^2(3, N = 299) = 11.50$, $p = .009$) with a medium effect size (Cramer's V = .20). Students in the summative SSA group were more often represented in the very deep approach profile, and less often represented in the low deep approach profile.

Finally, three of the larger student profiles were investigated regarding the study variables between the two SSA models. First, there were no differences between the reported means of surface and deep approach in any of the profiles, other than the almost significant difference in the deep approach profile. Within the deep approach profile, students in the summative SSA group reported a slightly larger amount ($t(115) = 1.91$, $p = .058$, $d = .36$) of deep approach to learning (M = 3.71, SD = .66) than the students in the formative SSA group (M = 3.47, SD = .69). However, greater differences were found regarding students' self-efficacy, which was reported as being higher in the summative SSA group in both very deep approach (M_{summ} = 4.74, SD_{summ} = .39; M_{form} = 4.37, SD_{form} = .48; $t(115) = 4.46$, $p < 0.001$, $d = .83$) and deep approach profiles (M_{summ} = 4.18, SD_{summ} = .50; M_{form} = 3.81, SD_{form} = .65; $t(115) = 3.32$, $p < 0.001$, $d = .64$). The effect sizes in both groups were large. In terms of course achievement, the student profiles were generally homogeneous. The only significant difference was found in the dissonant approach profile, in which the students in the summative SSA group scored significantly lower (M_{summ} = .66, SD_{summ} = .15; M_{form} = .74, SD_{form} = .12; $t(51) = .77$, 9 < 0.05, $d = .58$). In summary, the profiles were generally coherent regarding the variables of the study. The most significant differences were identified regarding self-efficacy in the two largest student profiles.

Discussion

The present study widens the literature on summative self-assessment in HE. Drawing on person-oriented analysis, summative and formative models of SSA were empirically compared in terms of students' approaches to learning, self-efficacy and course achievement.

Overall, the profile analysis showed that students in both SSA groups applied high levels of the deep approach. This is unusual, since the context of science has earlier been related to high levels of the surface approach (Parpala et al. 2010); the student-centred learning environment implemented in both SSA groups might be the reason behind this. Also, a link between the deep approach to learning and higher course achievement was found, which is in line with previous research (Diseth 2003; Sadler and Good 2006). Interestingly, within the student profile applying both the deep and surface approaches (*dissonant profile*), students in the summative SSA group scored lower in achievement than students in the formative group. This might imply that some students who would usually apply the surface approach in their studying, might not be able to adapt easily to summative SSA which favours the deep approach (e.g. Marton and Säljö 1976).

Although all the students showed a surprisingly high level of the deep approach, both general-level and person-oriented analyses revealed that the summative SSA model was able to promote the deep approach more than the formative one. Earlier studies (e.g. Baeten et al. 2008) have found that alternative assessment can be used to prevent passive learning. Here, a profile analysis showed that summative SSA did not exactly discourage the surface approach, but it did support the deep approach. Previously, it has been questioned whether the deep approach can be ‘induced’ with assessment (Haggis 2003; Struyven et al. 2006) and that alternative assessment might even lead to an increase in the surface approach (Gijbels and Dochy 2006). What features of summative SSA made this possible, since similar results are rarely reported? While the present quantitative study cannot directly answer this question, some hypotheses can be drawn up. As Sadler and Good (2006) highlighted, self-assessment might enhance a deeper understanding of the content if it is truly aligned with the pedagogical purposes of education. We argue that our implementation of the summative SSA model was perceived by the students as future-driven and as aligned with the purpose of life-long learning (Boud and Falchikov 2006; Tan, 2007, 2009). We hypothesise that self-grading was needed to foster the idea that self-assessment is done for the students *themselves*, not for the teacher. Thus, summative self-assessment might have led to different kind of student agency than formative self-assessment (Taras 2015).

Our results show substantial differences between the two SSA models regarding self-efficacy beliefs. The summative SSA model was largely connected with higher levels of self-efficacy. Interestingly, in the *very deep* and *deep* approach profiles, the students' mathematical achievement did not differ between the SSA groups, but their self-efficacy substantially did. This might be due to giving students more power over their assessment (Taras 2015, 2008) leading to different kinds of learner agency (Tan 2009, 2007). As Bandura (1997) suggested, students with strong self-efficacy set higher goals for themselves - perhaps the students in the summative SSA group were able to set goals for *themselves*, rather than studying for the exam. These results were found even though the digital validation system was used to check the final self-graded marks. It might even be that digital feedback supported students' beliefs of being capable of assessing themselves. Future research should draw on deeper data (e.g. interviews) to understand better the relationship between summative SSA and self-efficacy and further, their interconnection with the deep approach to learning, since our results show that a higher level of self-efficacy was connected with a greater level of the deep approach (see Diseth 2011; Prat-sala and Redford 2010). A deeper investigation of the notion of student *agency* might offer a key to understand these interrelations.

It is not enough to state that self-grading should not be used without offering empirical evidence (see Andrade and Cizek 2010; Andrade and Du 2007; Bourke 2018). Here, summative self-assessment was empirically shown to be able to support students' studying. We argue that the differences found between the SSA groups were based on the thorough implementation of the summative self-assessment model. However, the summative SSA model requires a context in which it can be substantially implemented. Thus, instead of investigating the ways in which SSA practices could be used, focus should be turned towards observing the educational contexts in which these practices are conducted. Future research could look for the characteristics of those cultures and learning environments that allow successful implementation of future-driven SSA (Tan 2007, 2009). As balancing between various purposes of assessment is complicated in HE (Broadbent et al. 2017), this offers a challenging task to both educators and researchers. Implementing only parts of future-driven SSA models might not be able to support studying in a desirable way, as our results on the formative SSA model imply.

The present study suggests that in our context, summative SSA could be implemented to align the purpose of assessment with the educational goals of HE. Effective use of summative SSA demands a conceptual change in what we mean by self-assessment, and this shift needs to be further transferred in pedagogical practices. Summative SSA challenges our usual norms of assessment, but given that we aim to foster meaningful study methods and lifelong learning in HE, is the idea of it all that radical?

Acknowledgements

The authors would like to thank Jokke Häsä from the DISA project for his great contribution as the inventor of the research design of this study. The authors would also like to express their gratitude to Jani Hannula, Juulia Lahdenperä, Saara Lehto and Jenni Räsänen for their comments on the manuscript and for their endless support. Also, the authors gratefully thank the referees for their input on improving the quality of the article.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Alonso-Tapia, Jesús, and Ernesto Panadero. 2010. "Effects of Self-Assessment Scripts on Self-Regulation and Learning." *Infancia y Aprendizaje* 33 (3): 385–397.

Andrade, Heidi, and G. J. Cizek. 2010. "Students as the Definitive Source of Formative Assessment: Academic Self-Assessment and the Self-Regulation of Learning."

In *Handbook of Formative Assessment*, edited by Heidi Andrade, and G. J. Cizek, 102–117. New York: Routledge.

Andrade, Heidi, & Ying Du. 2007. “Student Responses to Criteria-Referenced Self-Assessment.” *Assessment & Evaluation in Higher Education* 32 (2): 159–181.

Akaike, Hirotugu. 1987. “Factor Analysis and AIC.” *Psychometrika* 52, 317–332.

Asikainen, Henna and David Gijbels. 2017. “Do Students Develop Towards More Deep Approaches to Learning During Studies? A Systematic Review on the Development of Students’ Deep and Surface Approaches to Learning in Higher Education.” *Educational Psychology Review* 29 (2): 205–234.

Baeten, Marlies, Filip Dochy, and Katrien Struyven. 2008. “Students’ Approaches to Learning and Assessment Preferences in a Portfolio-Based Learning Environment.” *Instructional Science* 36 (5–6): 359–374.

Bandura, Albert. 1997. *Self-Efficacy*: *The Exercise of Control*. New York: Freeman.

Beaumont, Chris, Michelle O’Doherty, and Lee Shannon. 2011. “Reconceptualising Assessment Feedback: A Key to Improving Student Learning?” *Studies in Higher Education* 36 (6): 671–687.

Boud, David, and Nancy Falchikov. 2006. “Aligning Assessment With Long-Term Learning.” *Assessment & Evaluation in Higher Education* 31 (4): 399–413.

Bourke, Roseanna. 2018. “Self-Assessment to Incite Learning in Higher Education: Developing Ontological Awareness.” *Assessment & Evaluation in Higher Education* 43 (5): 827–839.

Broadbent, Jaclyn, Ernesto Panadero, and David Boud (2018). “Implementing Summative Assessment With a Formative Flavour: A Case Study in a Large Class.” *Assessment & Evaluation in Higher Education* 43 (2), 307–322.

Brown, G. T., H. L. Andrade, and Fei Chen. 2015. “Accuracy in Student Self-Assessment: Directions and Cautions for Research.” *Assessment in Education: Principles, Policy & Practice* 22 (4), 444–457.

Brown, Gavin, and L. R. Harris. 2013. “Student Self-Assessment.” In *The Sage Handbook of Research on Classroom Assessment*, edited by J. H. McMillan, 367–393. Thousand Oaks, CA: Sage.

Collins, L. M., and S. T. Lanza. 2010. Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences. Hoboken, NJ: John Wiley & Sons.

Crisp, G. T. 2012. ”Integrative Assessment: Reframing Assessment Practice for Current and Future Learning.” *Assessment & Evaluation in Higher Education* 37 (1): 33–43.

Diseth, Åge. 2003. “Personality and Approaches to Learning as Predictors of Academic Achievement.” *European Journal of Personality* 17 (2): 143–155.

Diseth, Åge. 2011. “Self-efficacy, Goal Orientations and Learning Strategies As Mediators Between Preceding and Subsequent Academic Achievement.” *Learning and Individual Differences* 21 (2): 191–195.

Entwistle, N. J., and Paul Ramsden. 1983. *Understanding Student Learning*. London: Croom Helm.

Entwistle, Noel. 2009. *Teaching for Understanding at University: Deep Approaches and Distinctive Ways of Thinking*. Basingstoke: Palgrave Macmillan.

Gijbels, David, and Filip Dochy. 2006. “Students’ Assessment Preferences and Approaches to Learning: Can Formative Assessment Make a Difference?” *Educational studies* 32 (4): 399–409.

Herrmann, K. J., Anna Bager-Elsbor, and Anna Parpala. 2017. “Measuring Perceptions of the Learning Environment and Approaches to Learning: Validation of the Learn Questionnaire.” *Scandinavian Journal of Educational Research,* 61 (5): 526–539.

Ibabe, Izaskun, & Jauregizar, Joana. 2010. Online Self-Assessment With Feedback and Metacognitive Knowledge. *Higher Education 59* (2): 243–258.

Haggis, Tamsin. 2003. Constructing Images of Ourselves? A Critical Investigation Into ‘Approaches to Learning’ Research in Higher Education. *British Educational Research Journal* 29 (1): 89–104.

Hu, Lit-ze, & P. M. Bentler. 1999. “Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives.” *Structural Equation Modeling* 6: 1–55.

Jay, Julie, & Owen, Antonette. 2016. “Providing Opportunities for Student Self-Assessment: The Impact on the Acquisition of Psychomotor Skills in Occupational Therapy Students.” *Assessment & Evaluation in Higher Education 41* (8): 1176–1192.

Knight, P. T. 2002. “Summative Assessment in Higher Education: Practices in Disarray.” *Studies in Higher Education* 27 (3): 275–286.

Levine, Arthur, and D. R. Dean. 2012. *Generation on a Tightrope: A Portrait of Today's College Student*. San Francisco: John Wiley & Sons.

Lindblom-Ylänne, Sari. 2003. Broadening Understanding of the Phenomenon of Dissonance. *Studies in Higher Education* 28 (1): 63–77.

Lo, Yungtai, N. R. Mendell, and D. B. Rubin. 2001. Testing the Number of Components in a Normal Mixture. *Biometrika* 88 (3): 767–778.

López-Pastor, V. M., J. M. Fernández-Balboa, M. L. Santos Pastor, and A. A. Fraile. 2012. "Students' Self-Grading, Professor's Grading and Negotiated Final Grading at Three University Programmes: Analysis of Reliability and Grade Difference Ranges and Tendencies." *Assessment & Evaluation in Higher Education* 37 (4): 453–464.

Marton, Ference, and Robert Säljö. 1976. "On Qualitative Differences in Learning: I—Outcome and Process." *British Journal of Educational Psychology* 46 (1): 4–11.

Panadero, Ernesto, Anders Jonsson, and Juan Botella. 2017. "Effects of Self-Assessment on Self-Regulated Learning and Self-Efficacy: Four Meta-Analyses." *Educational Research Review 22*: 74–98.

Panadero, Ernesto, G. T. Brown, and J. W. Strijbos. 2016. "The Future of Student Self-Assessment: A Review of Known Unknowns and Potential Directions." *Educational Psychology Review* 28 (4): 803–830.

Panadero, Ernesto, J. A. Tapia, and J. A. Huertas. 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.

Panadero, Ernesto, and Margarida Romero. 2014. "To Rubric Or Not to Rubric? The Effects of Self-assessment on Self-regulation, Performance and Self-efficacy." *Assessment in Education: Principles, Policy & Practice* 22 (2): 133–148.

Parpala, Anna, and Sari Lindblom-Ylänne. 2012. "Using a Research Instrument for Developing Quality at the University." *Quality in Higher Education* 18 (3): 313–328.

Parpala, Anna, Sari Lindblom-Ylänne, Erkki Komulainen, Topi Litmanen, and Laura Hirsto. 2010. "Students' Approaches to Learning and Their Experiences of the Teaching–Learning Environment in Different Disciplines." *British Journal of Educational Psychology* 80 (2): 269–282.

Pintrich, P. R., D. A. F. Smith, Teresa Garcia, and W. J. McKeachie. 1991. *A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ)*.

Ann Arbor, MI: National Center for Research to Improve Post-secondary Teaching.

Postareff, Liisa, Viivi Virtanen, Nina Katajavuori, and Sari Lindblom-Ylänne. 2012. “Academics’ Conceptions of Assessment and Their Assessment Practices.” *Studies in Educational Evaluation* 38 (3–4): 84–92.

Prat-Sala, Mercè, and Paul Redford. 2010. The Interplay Between Motivation, Self-efficacy, And Approaches to Studying. *British Journal of Educational Psychology 80* (2): 283–305.

Richardson, Michelle, Charles Abraham, and Rod Bond. 2012. “Psychological Correlates of University Students’ Academic Performance: A Systematic Review and Meta-Analysis.” *Psychological Bulleting* 138 (2): 353–387.

Rust, Chris, Berry O’Donovan, and Margaret Price. 2005. “A Social Constructivist Assessment Process Model: How the Research Literature Shows Us This Could Be Best Practice.” *Assessment & Evaluation in Higher Education* 30 (3): 231–240.

Rämö, Johanna, Reinholz, Daniel, Häsä, Jokke, & Lahdenperä, Juulia. 2019. “Extreme Apprenticeship: Instructional Change as a Gateway to Systemic Improvement.” *Innovative Higher Education*, 1-15. https://doi.org/10.1007/s10755-019-9467-1

Sadler, P. M., and Eddie Good. 2006. “The Impact of Self- and Peer-Grading on Student Learning.” *Educational Assessment* 11 (1): 1–31.

Shute, V. J., and Y. J. Kim. 2014. “Formative and Stealth Assessment.” In *Handbook of Research on Educational Communications and Technology*, edited by J.M Spector, M.D. Merrill, J. Elen and M. J. Bishop, 311–323. New York: Lawrence Erlbaum Associates.

Schwarz, Gideon. 1978. “Estimating the Dimension of a Model.” *The Annals of Statistics*, 6 (2): 461–464.

Strong, Brent, Mark Davis, and Val Hawks. 2004. “Self-Grading in Large General Education Classes: A Case Study.” *College Teaching* 52 (2): 52–57.

Struyven, Katrien, Filip Dochy, Els Janssens, and Sarah Gielen. 2006. “On the Dynamics of Students' Approaches to Learning: The Effects of the Teaching/Learning Environment.” *Learning and Instruction* 16 (4): 279–294.

Tan, Kelvin. 2007. “Conceptions of Self-Assessment: What is Needed for Long Term Learning?” In *Rethinking Assessment in Higher Education: Learning for the*

Longer Term, edited by David Boud, and Nancy Falchikov, 114–127. London: Routledge.

Tan, K. H. 2009. "Meanings and Practices of Power in Academics' Conceptions of Student Self-Assessment." *Teaching in Higher Education* 14 (4): 361–373.

Taras, Maddalena. 2008. "Issues of Power and Equity in Two Models of Self-Assessment." *Teaching in Higher Education* 13 (1): 81–92.

Taras, Maddalena. 2015. "Situating Power Potentials and Dynamics of Learners and Tutors Within Self-Assessment Models." *Journal of Further and Higher Education* 40 (6): 846–863.

Tejeiro, R. A., J. L. Gómez-Vallecillo, A. F. Romero, Manuel Pelegrina, Agustín Wallace, and Enrique Emberley. 2012. "Summative Self-Assessment in Higher Education: Implications of Its Counting Towards the Final Mark." *Electronic Journal of Research in Educational Psychology* 10 (2): 789–812.

Universities Act 558/2009. Issued 24 July 2009 in Helsinki, Finland.

Tables

Table 1. Part of the rubric of the course. Each topic was divided into three sections (skills corresponding to grades 1-2, 3-4 and 5) and consisted of multiple learning objectives.

	Skills corresponding to grades		
Topic	1-2	3-4	5
Matrices	I can perform basic matrix operations and know what zero and identity matrices are	I can check, using the definition of an inverse, whether two given matrices are each other's inverses	I can apply matrix multiplication and properties of matrices in modelling practical problems
Reading and writing	I use course's notation in my answers	In my solutions, I write complete, intelligible sentences that are readable to others	I can write proofs for claims that concern abstract or general objects

Table 2. Participants of the study.

		Summative SSA group N = 152		Formative SSA group N = 147		Total N = 299	
Descriptives		N	%	N	%	N	%
Major	Mathematics	48	31.6	45	30.6	93	31.1
	Related science	57	37.5	59	40.1	116	38.8
	Other	47	30.9	43	29.2	90	30.1
Gender	Female	52	34.2	51	34.7	103	34.4
	Male	94	61.8	92	62.6	186	62.2
	Other / I don't want to answer	6	4.0	4	2.7	10	3.3
Achievement on the course		Mean	SD	Mean	SD	Mean	SD
Tasks with	**automatic feedback** (max. 70 scores)	51.41	7.63	52.13	6.59	51.76	7.14
	teacher feedback (max. 10 scores)	8.72	1.44	8.59	1.34	8.66	1.39
	no feedback (max. 53 scores)	37.18	13.70	38.02	12.15	37.59	12.95
	Achievement score (max. 1 score)	.77	0.14	.77	.12	.77	.13

Table 3. Approaches to learning and self-efficacy in the two self-assessment groups.

		Surface approach		Deep approach		Self-efficacy	
	N	Mean	SD	Mean	SD	Mean	SD
Summative SSA group	152	1.93	.79	3.84	.72	4.28	.74
Formative SSA group	147	2.16	.78	3.56	.80	3.82	.83
Total	299	2.04	.8	3.7	.77	4.05	.82

Table 4. Fit indices for the profile solutions.

	2 profiles	3 profiles	4 profiles	5 profiles	6 profiles
AIC	1370.584	1353.499	1337.643	1331.445	1319.864
BIC	1396.487	1390.503	1385.748	1390.653	1390.173
aBIC	1374.287	1358.789	1344.52	1339.91	1329.916
VLMR	-700.297	–678.292	-666.749	-655.821	-649.723
p_{VLMR}	0.0009	0.1424	0.0238	0.0094	0.1162
LMR LRT	41.579	21.81	20.648	11.523	16.61
$p_{\text{LMR LRT}}$	0.0013	0.1557	0.0279	0.0118	0.1324
Smallest profile (%)	14.72	5.35	5.02	0.33	1.67

Table 5. ANOVA comparison between the student profiles.

	Profile 1 Deep approach (N = 116)		**Profile 2** Low deep approach (N = 116)		**Profile 3** Dissonant approach (N = 52)		**Profile 4** Surface approach (N = 15)		ANOVA		Post hoc (Tukey HSD) comparison
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	$F(4, 298)$	η^2	
Deep approach	4.07	0.67	3.57	0.69	3.34	0.78	3.12	0.79	20.01*	0.17	1 > 2,3,4
Surface approach	1.28	0.26	2.14	0.24	2.93	0.26	4.04	0.35	834.00*	0.89	4 > 1,2,3; 3 > 1,2; 2 > 1
Self-efficacy	4.60	0.46	3.96	0.62	3.41	0.87	2.79	0.88	68.13*	0.41	1 > 2,3,4; 2 > 3,4; 3 > 4
Achievement on the course	0.82	0.12	0.77	0.11	0.70	0.14	0.66	0.13	15.72*	0.14	1 > 2,3,4; 2 > 3,4

* $p < 0.001$

Figures

Figure 1. An overview of the design of the study. The summative and formative models only differed in terms of their final, summative grading method.

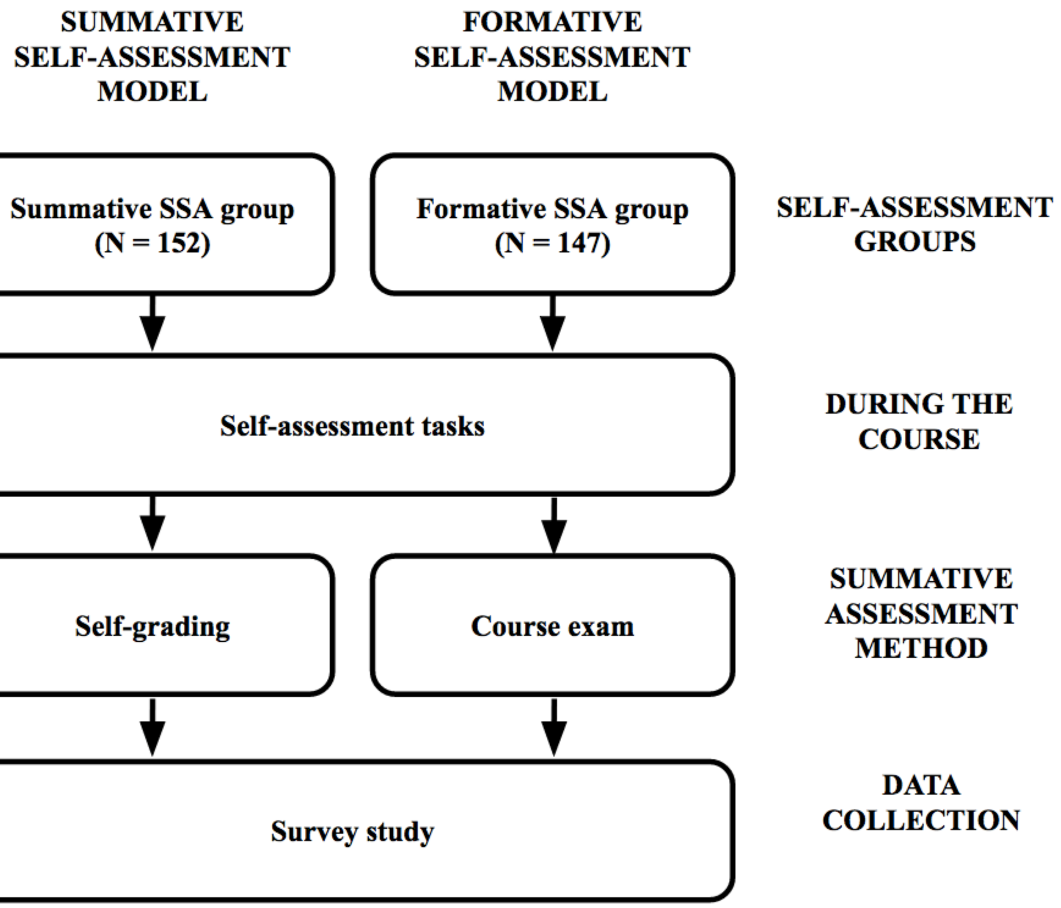

Figure 2. Z-scores of the variables of the study of the four student profiles.

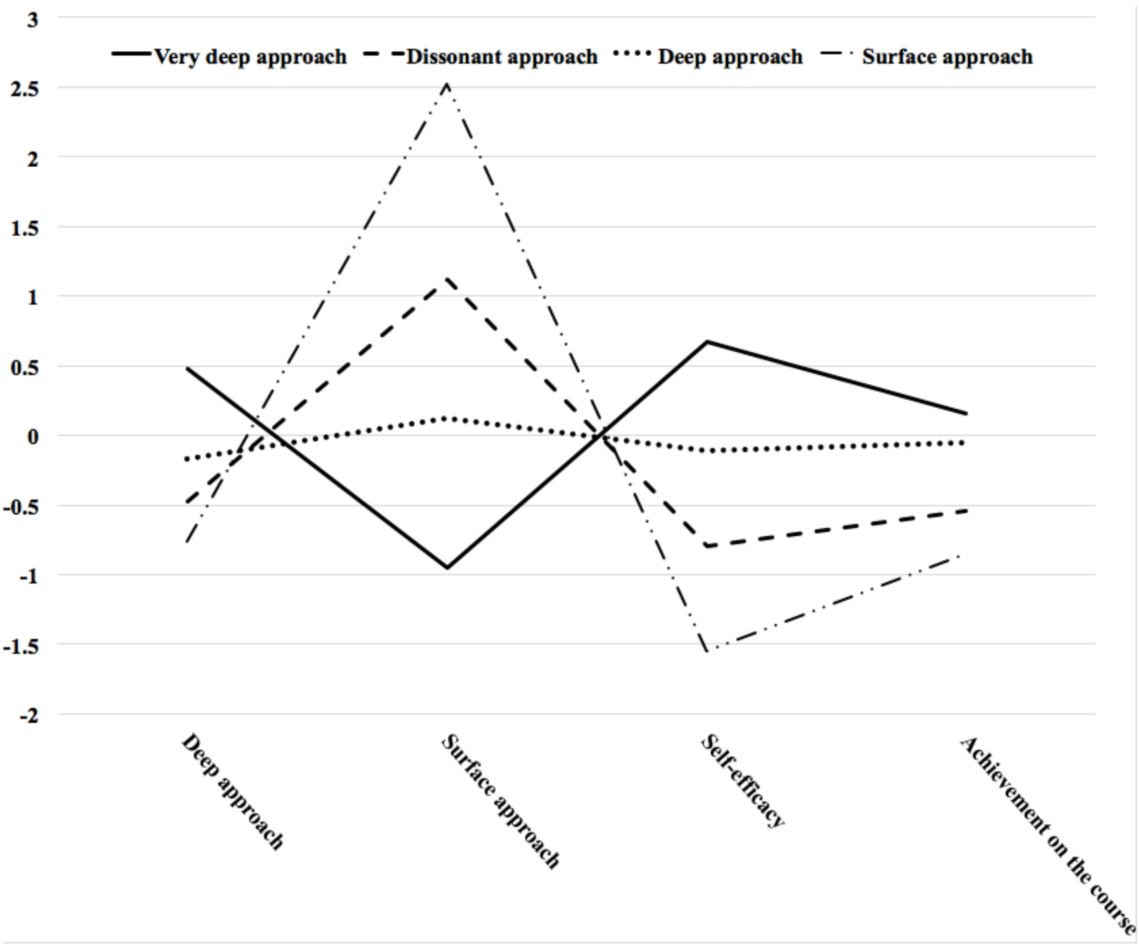

Figure 3. The distribution of the student profiles in the two SSA models.