



An assessment innovation as flywheel for changing teaching and learning

Judith T. M. Gulikers, Piety Runhaar & Martin Mulder

To cite this article: Judith T. M. Gulikers, Piety Runhaar & Martin Mulder (2018) An assessment innovation as flywheel for changing teaching and learning, Journal of Vocational Education & Training, 70:2, 212-231, DOI: [10.1080/13636820.2017.1394353](https://doi.org/10.1080/13636820.2017.1394353)

To link to this article: <https://doi.org/10.1080/13636820.2017.1394353>



© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 27 Oct 2017.



Submit your article to this journal [↗](#)



Article views: 1511



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 2 View citing articles [↗](#)

An assessment innovation as flywheel for changing teaching and learning

Judith T. M. Gulikers, Piety Runhaar and Martin Mulder

Education and Competence Studies, Social Sciences Group, Wageningen University and Research, Wageningen, The Netherlands

ABSTRACT

This contribution shows an example of how an assessment innovation can serve as a flywheel for changing teaching and learning. This article first explains the development of an authentic, competence-based performance assessment in pre-vocational secondary education (pre-VET). Using these assessments in *pre-VET*, that is preparatory to VET, is important for motivating students for a future in VET. Second, in an action research 34 teacher teams of 11 pre-VET institutions implement the new assessment, supported by researchers and teacher coaches, to illuminate how it motivates students, prepares them for VET, and influences teaching and learning. Data from 76 teachers, 68 students, 24 teacher coaches and 3 researchers showed how implementing these assessments raised questions about changing educational content and pedagogy to fit a competence-based approach. Increasing teachers' expectations of students, changing student-teacher interaction patterns, and developing teachers' practical understanding of competencies and how they can be observed and discussed in assessments are some of the intriguing challenges that require further support for proper implementation of competence-based assessments and developing an aligned competence-based teaching and learning process.

ARTICLE HISTORY

Received 16 June 2016
Accepted 21 September 2017

KEYWORDS

Assessment innovation; competence-based assessment; competence-based teaching; pedagogy; teacher professional development

Introduction

Assessment steers teaching and student learning. If we want to change what and how students learn and what and how teachers teach, we need to change the assessments (Popham 2001). This assumption lies at the heart of the assessment innovation discussed in this article.

Many countries have enacted outcome-based qualification frameworks (Young 2009), like the European Qualifications Framework (EQF), defining relevant

CONTACT Judith T. M. Gulikers  judith.gulikers@wur.nl

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

outcomes that students should develop during their educational programme. These outcomes guide curriculum development, teaching, learning and assessment. Certainly in Vocational Education and Training (VET), many education reforms have taken place in pedagogy and curriculum to increase student outcomes, in terms of knowledge, skills or competencies, relevant for the future workplace (Weigel, Mulder, and Collins 2007).

However, Achtenhagen and Winther (2014) argue that (VET) research largely ignores the assessment side of these educational innovations. Looking from the perspective of the curriculum-instruction-assessment triad (Achtenhagen 2012), or Biggs (1996) idea of constructive alignment between teaching, learning and assessment, a curriculum innovation cannot succeed if the assessment side of the triad is not seriously taken into account. The assessment innovation described in this article, even goes a step further by arguing that a curriculum innovation should *start* with changing the assessment.

The article is divided in two parts. Part one elaborates on the development and theoretical underlying rationales of a new authentic, competence-based performance assessments in Dutch pre-vocational secondary education (pre-VET), that is preparatory for VET. Part two is a large scale action research of 34 teacher teams of 11 pre-VET institutions implementing the new assessment. The study shows how this assessment influences student motivation and preparation for VET. Moreover, it illuminates what additional changes appear to be needed in teaching content and pedagogy to fit the competence-based approach (Sturing et al. 2011).

This study can be exemplary for other (pre-)VET institutions thinking about or struggling with developing and implementing authentic, competence-based performance assessment, teaching and learning. It shows the challenges that teachers and students face when changing from 'traditional' teaching and learning to competence-based teaching and learning. Despite the increased body of *theoretical* knowledge on the topic, empirical studies and, thus, practical knowledge is still scarce (Dietzen and Nicolaus 2016)

Research context

Pre-vocational education (pre-VET) precedes vocational education (VET) and is a type of secondary education for 12-16 year old students with more practical, rather than theoretical and orientation. Their learning preference is largely learning by doing and they are often focused on learning for an occupation. The motivation of pre-VET students is a long known and tough phenomenon in the Netherlands (Inspectorate of Education 2014) resulting in early drop-outs or inappropriate study choices for future VET trajectories.

Pre-VET focuses on EQF level 3 (see Table 1) and 40% of all Dutch pupils are enrolled in this type of education. It gives students an orientation on and preparation for a broad variety of vocational fields for which they can choose a specific programme in VET. Broadly speaking, these fields are (health) care, technology, trade/

Table 1. Learning outcomes of EQF Level 3.

	Knowledge	Skills	Competencies
The learning outcomes relevant to EQF Level 3 are	Knowledge of facts, principles, processes and general concepts, in a field of work or study	A range of cognitive and practical skills required to accomplish tasks and solve problems by selecting and applying basic methods, tools, materials and information	Take responsibility for completion of tasks in work or study; adapt own behaviour to circumstances in solving problems

administration/economy and agrifood. This study is aimed at the latter field, which is also referred to as 'green education'. This field goes beyond primary production and includes an orientation on plant and production studies, design and creative studies, living environment and leisure, food and health and animal and care.

Assessment in pre-VET: current practices

Graduating from pre-VET comprises a national practical and written exam at the end of the final (i.e. fourth) year (50%) and a combination of classroom assessments (50%) during year three and four. Schools and teachers can decide how many and what kind of assessments they combine in their classroom assessments, as long as it contains theoretical and practical tests. Moreover, classroom assessments should be targeted at assessing the learning outcomes of EQF level 3 (see Table 1, right column) as this prepares students for VET (EQF level 4). Current classroom assessment practices can be characterised as follows: theory assessments of closed or open answer formats focusing mostly on low cognitive levels of Bloom. The practical tests focus on isolated testing of routine and/or technical skills like syringe putting, potting plants or VAT calculations. The competencies column of Table 1 is largely ignored.

Assessment in pre-VET: the assessment innovation project 'Green Assessments'

The classroom assessment innovation project described in this article was named 'Green Assessments' (i.e. performance assessments in green education) and focuses on assessing these competencies. It developed authentic performance assessments of competencies from the vision (1) that assessment steers learning; (2) that the role of assessment in pre-VET as a preparatory programme should be motivational, and (3) offer young students a situated preview of possible occupational practices of follow-up VET trajectories, and (4) that the assessments should offer students the opportunity to experience their own competence in relation to these occupational practices. Thus, the project adopted an 'assessment *for* learning' focus.

The ultimate aims of project were to (1) stimulate a transition towards more competence-based teaching and learning in pre-VET, and (2) help building a continuous learning pathway from pre-VET green education to competence-based green VET that is already working with these kinds of authentic, competence-based performance assessment.

PART I: the development of the authentic, competence-based performance assessments

This section describes the theoretical rationales with respect to conceptualisation of competence and assessment of competencies and elaborates on how these theoretical notions are translated into the assessment design.

Conceptualisations of competence

This study adheres to the 'pyramid' of Miller (1990) of professional competence which is often used in Dutch educational programmes and assessment research (e.g. Baartman and Gulikers 2017). Professional competence means the capability to properly perform occupational tasks and jobs in their realistic complexity (Miller 1990). Miller views professional competence as built-up of four layers, being 'knows', 'knows how', 'shows how' and 'does' in which the lower levels are prerequisite for the higher levels, while being competent means being able to perform at the highest (does) level. This highest level requires an integration of knowledge, skills and attitudes in competent performance, assuming that competent performance cannot be done when underlying knowledge and skills are not internalised. To assess if a student is competent, the assessment should address the does-level. This pyramid makes a direct link between the 'what' and the 'how' of assessment (Baartman and Gulikers 2017). At the top of the pyramid, assessment methods become increasingly authentic and complex, meaning that the assessment increasingly resembles professional practice at the level of the students (Gulikers, Bastiaens, and Kirschner 2004)

Additionally, competence is conceptualised as *situated professionalism* (Mulder 2014). Mulder distinguishes three approaches of competence and elaborates on how these have guided the development of occupational and practice-based learning programmes. *First*, behaviouristic functionalism in which competence is subdivided in a range of measurable behaviours, which often resulted in minuscule and fragmented skills lists, regularly ignoring how these should be integrated in the action repertoire of graduates. *Second*, integrated occupationalism, in which competence is viewed as a set of *competencies*, which are integrations of knowledge, skills and attitudes, required to deal with specific occupational roles and tasks. This approach is mainly industry-led. The defined competencies for a competence-based educational programme are determined by the requirements of a certain profession. And *third*, situated professionalism, in which competence gets its meaning in the specific context in which it is used. This approach focused on generic competencies relevant for many professions, like planning or communicating, but stresses the importance of giving them meaning in a certain context. For example, communicating for a teacher means something completely different than communicating for a researcher or a doctor. Teaching and assessments should

operationalise the meaning of the generic competencies within this specific situation (Wesselink et al. 2017).

Although most competency-based frameworks of contemporary VET programmes are grounded in the integrated occupationalism approach (Mulder 2014), the situated professionalism approach fits the Dutch pre-VET context better. Compared to VET, this context is less industry-led, has an orientation function and aims at fostering the development of generic competencies relevant in a wide array of future jobs.

Conceptualisation of assessing competencies

Assessing competence in the sense of assessing the students' capability to integrate knowledge, skills and attitudes at Miller's does-level in authentic occupational tasks, obviously requires a different kind of assessment than knowledge-based paper-and-pencil tests. This has led to several studies on what competence-based assessments should look like (e.g. Gulikers, Biemans, and Mulder 2009; Sturgis 2014). This study embraces the ideas of Competence Assessment Programmes (Baartman et al. 2006). A core underlying idea is that competencies cannot be assessed using one assessment method at one moment in time, but that a combination of assessments – a competence assessment *programme* (CAP) – is required. Such an assessment programme refers to a collection of assessments, in which traditional (e.g. paper-and-pencil test) and new (e.g. performance assessments) assessments; assessment of products as well as processes; and formative as well as summative assessment can be combined to allow for a qualitatively good judgement of students' competencies. As a guideline, Baartman and Gulikers (2017) say that an assessment programme should at least combine an authentic does/shows how-level assessment with a knows/knows how-level assessment. This line of thinking requires a different view on assessment quality, than a strict psychometric perspective (Baartman et al. 2006; Knight 2000). Psychometric quality criteria (e.g. reliability and validity) should be seen in a different light and edumetric quality criteria (e.g. authenticity, educational consequences) gain importance. For example, the quality criterion of *reliability* in the sense of objectivity and standardisation is put in a different perspective, namely that of *reproducibility and comparability*. That means that purposeful actions should be put in place to make a more subjective assessment (e.g. an observation of performances) reproducible between different assessors and comparable between different students, without requiring the assessment should be exactly the same. For example, by providing assessors observation schemes or undertaking moderation and calibration sessions for more comparable grading practices (Sturgis 2014). This view on competence-based assessment highly values the teacher/assessor as a professional (McMullan et al. 2003) and entails that assessment quality depends both on the assessment instrument as well as on how it is implemented by teachers in educational practice (Hawkins et al. 2015).

An additional new perspective that this CAP thinking brings to the assessment debate is that a qualitatively good assessment programme does not only have a summative function. It should also positively influence teaching and learning (i.e. a formative assessment). For this, a close alignment between assessment–teaching–learning is prerequisite (Achtenhagen and Winther 2014).

Implications of theory for the assessment design

The adopted conceptualisations of competence and assessment of competencies have implications for the development of the new assessment. Theoretical rationales underlying the new assessment are (1) the assessment assesses the integration of knowledge, skills and attitudes at the does-level of Millers pyramid, which (2) requires an authentic assessment of real-life, complex practical performance at the level of pre-VET students in which (3) generic competencies should be given a concrete meaning. Additionally, (4) it should combine at least two assessment methods of which one focuses on Millers’ does/show how level and one on Millers’ knows/knows how level; (5) various actions should be put in place in the assessment design to help assessors to evaluate students’ competencies as objective as possible, while at the same time, not trivialising the assessment into ticking off analytic criteria (Harsch and Martin 2012), and (6) the assessments should be embedded in a competence-based teaching and learning process.

The next section elaborates on the characteristics of the new assessment, followed by Part II that describes the action research studying the effects that proper implementation of these new assessments has on teaching and learning.

The product: the design of the authentic competence-based performance assessment

Representatives of 11 participating green pre-VET institutions (from the 12 green pre-VET schools in total in the Netherlands), both managers and teachers, joined forces in the assessment innovation project Green Assessments. Two project leaders, both assessment experts of the council for green education, guided this project. It started by explicating an assessment vision using the above mentioned theoretical rationales. This vision was translated in steps to follow in designing the new authentic performance assessments for the five green education fields mentioned above. Thus, every assessment followed the same structure and procedures (see below), while the concrete content of the assessments differed. This content was written by groups of representative teachers from the 11 pre-VET schools with relevant content knowledge (i.e. a group for developing assessments for floristry, another group for animal studies, etc.,). The assessment design steps, including the most important assessments characteristics, are the following (see also Figure 1):

Critical occupational tasks	Creating, packing, presenting, selling and advising about a range of floristry products
Authentic assessment task	“You work in Flower Shop “Hello Sunshine” together with your boss and two colleague florists. Miss Humphry, working at the geriatric home close to the shop, comes to the shop to discuss with you the possibilities of creating three flower arrangements to cheer up the diner room of the house. The arrangements should be lasting, bright in colours and without strong fragrance, and within the budget of 150 euros. She is an insecure lady asking you for advice and hopes you can help her make a decision that will make all the elderly patients in her house happy.”
Four generic competencies	Collaborating, communicating/presenting, showing professional expertise (knowledge and skills), and customer-oriented working
Operationalisation of the four generic competencies in the context of this assessment	<p><u>Collaborating</u> The student effectively discusses with boss, colleagues, and customer(s) about work to be done. The students shows flexibility in this collaboration process The student adjusts his/her work based on the discussed agreements.</p> <p><u>Communicating and presenting, both verbally and non-verbally</u> The students takes initiative in coming up with ideas The student provides clear and correct advice to customers The student pays attention to the verbal and nonverbal communication of the customers the students presents flower arrangements attractively, taking into account the characteristics of the flowers</p> <p><u>Professional expertise</u> The students’ performance shows proper knowledge of colour and shape of the flowers he/she is using The students’ performance shows understanding of the material and tools he/she is using and makes proper choices with respect to the material and tools used. The student wraps up the bouquets in an attractive and creative way The student can fill in the order form including VAT and cost calculations</p> <p><u>customer-oriented working</u> The student shows a friendly and appropriate attitude towards the customers The student identifies the customers’ wishes, advises the customer and performs the request neatly and correctly The student is able to deal with “difficult” customers</p>
Prerequisite knowledge, skills and attitudes	<p><u>Knowledge</u> cost calculations VAT calculations design of floristry products the most common products, tools and materials in a flower shop</p> <p><u>Skills</u> draw up invoices Helping customers displaying flower arrangements creating flower arrangements packing flowers using the cash register</p> <p><u>Attitudes</u> being customer-friendly show an open, sincere and respectful attitude</p>

Figure 1. Core assessment characteristics operationalised in a floristry assessment example.

- (1) First, critical occupational tasks representative for the five occupational field are identified (Achtenhagen and Winther 2014) by groups of teachers and professionals working in the field. These tasks, the expected competencies and the expected level of performance are laid down in qualification frameworks and as such describe what tasks graduates are expected to do. For example, in the floristry field these are: creating, packing, presenting, selling and advising about a range of floristry products. In the field of animal studies these are: taking care of and

feeding a range of animals, taking care of animal housing, and guiding animal birth and breeding.

- (2) The actual assessment development starts with writing an authentic case description relating to one or more critical occupational tasks. This case description should require students to integrate knowledge, skills and attitudes into performing an real occupational whole task instead of a demonstration of isolated skills (Van Merriënboer 1997). The task should always contain a request of a client, result in a real authentic product or service, and have realistic complexity *at the student-level* (Gulikers, Bastiaens, and Kirschner 2004, 2006). Obviously, pre-VET students cannot perform real occupational tasks as graduated employees would, however, Gulikers, Bastiaens, and Kirschner (2004) argue that authentic assessment of professional task performance is possible at every educational level by asking the question: 'What kind of tasks would these students be able to do when doing an internship or performing this task in real life?' The complexity of the assessment task should resemble the answer to this question. A short version of an authentic case for 'floristry' students is shown in Figure 1.
- (3) For every authentic case, four generic competencies – out of 18 nationally defined competencies for VET and pre-VET like 'decision-making' or 'customer-oriented working' – are selected to be assessed. These four competencies are described in terms of concrete observable behaviour that is expected from students within their assessment task. Figure 1 shows the four competencies assessed in the floristry example and how these generic competencies are given concrete meaning in the situated task of the assessment.
- (4) Underlying required knowledge, skills and attitudes are mentioned in the assessment as prerequisites (see Figure 1) and not explicitly assessed. Preparatory education should have paid attention to these knowledge, skills and attitude aspects.
- (5) Each assessment contains a planning-, a performance- and a reflection phase: Students make a planning using a standard planning form, perform the tasks during which their performance is observed by two assessors, and justify their performance and choices in an assessment interview with the assessors afterwards. As such, this assessment contains three assessment methods that in combination lead to a judgement of students' competencies.
- (6) A standard planning form, an observation form containing the operationalisation of the four competencies that need to be observed (see Figure 1) and guidelines for the assessment interview are put in place to increase the comparability and reproducibility of the assessments (Baartman et al. 2006). That is, these help assessors to assess all students

in the same way and evaluate their competencies as objectively as possible.

- (7) Every assessment takes about 4 hours in total. Four hours observation combined with an assessment interview is expected to be adequate for assessing competencies in their full complexity.
- (8) Students are assessed by two assessors: the student's own teacher and another teacher from the same school or from another pre-VET school also working with the new assessment. They observe student performance and conduct the assessment interview.
- (9) Every assessment uses the same standard assessment scoring form to evaluate students' performance. This scoring form is exactly the same for all assessments, independent of their content, and combines holistic with analytic scoring (e.g. Harsch and Martin 2012). It starts with a holistic judgement: 'is the student competent in performing the professional whole task?' to be answered with yes or no. Then, this holistic evaluation is supported by more analytic scoring. Assessors score the following elements: the student's planning (1–10 points); the quality of the process in terms of student's performance of the four competencies operationalised in observable behaviour (see Figure 1; 1–40 points, 10 per competency); the quality of the result being either a product or provided service (1–30 points); and the student's justification of his/her performance of the four competencies within the whole tasks during the assessment interview (1–20 points). This is first done by both assessors individually, after which they compare their given scores and collaboratively agree on the scores given per element and the resulting total number of points (1–100). The number is translated into a final grade by dividing it by 10, leading to a final grade between 1 and 10. Additionally, the assessors add qualitative feedback in terms of *tops* and *tips*. Using the same assessment scoring form and procedure with two assessors is put in place to increase objective and fair evaluation of students' competencies and overall performance.
- (10) A structured and standard teacher guideline is provided that takes teachers through all the steps they need to take to get an assessment organised in their school. This ensures transparency for teachers/assessors and the comparability of the assessments between schools and enhances the likelihood that the assessments are implemented as intended.
- (11) For every assessment a student version is developed describing all relevant information that students need to know about this assessment, like what the assessment scoring form looks like, what competencies are assessed, in what context, what prerequisite knowledge and skills they need to possess, by whom they are assessed, and how long the

assessment will take. This increased the transparency for and comparability between students.

Quality check of the assessment product

Thirty-five new assessments are developed for the five fields of green (pre-)VET education. Every assessment is quality checked by an independent group of assessment experts from the Dutch national testing and assessment organisation (CITO) responsible for all national exams in primary and secondary education in the Netherlands. They evaluate the developed assessments using a framework of twelve quality criteria appropriate for competence-based assessments (Baartman et al. 2006).

Part II: studying the implementation and flywheel function

A large scale action research was set-up to experiment with implementing the new assessments for two reasons: (1) to help teacher teams to implement the assessments as intended and thereby increase the quality of the assessments (De Jonge et al. 2017; Gulikers et al. 2013), and (2) to develop an understanding of what implementing these competence-based assessments means for teaching content and pedagogy. Both implementing the assessments as intended as well as identifying required changes in the teaching and learning process are crucial for a transition to more competence-based pre-VET (Gulikers et al. 2013). Thirty-four teacher teams from 11 green pre-VET institutions participated voluntarily, for they wanted to innovate their curricula, teaching and learning towards competence-based education in alignment with competence-based green VET. Additionally, they believed that the new assessments would be more motivating for their students as they were more practical and situated in occupational contexts of possible future VET trajectories than the assessments they currently used. Next to the 34 teacher teams, one university researcher, the two project leaders of the Green Assessments Project, and 24 pre-VET teachers, who took part in developing the new assessments in Part I, participated in the action research to coach and support the teacher teams.

Research question

This research aims to shed a light on if, how and why the new competence-based performance assessment stimulates student learning, motivation and preparation for competence-based VET and how its implementation can be a flywheel for changing to more competence-based teaching and learning. Therefore, the research questions guiding this study are:

- (1) How do different stakeholder groups experience the effects of the new assessment on student

- (a) motivation;
 - (b) learning;
 - (c) preparation for VET, and
- (2) what characteristics of the implemented assessment led to these experienced effects.
 - (3) what do different stakeholder groups report regarding required changes in the teaching and learning process to develop aligned, competency-based practices?

Method

As said, 34 teacher teams of at least two teachers and a manager participated voluntarily. They all chose one new assessment from the field of their choice (e.g. floristry, plant and production) to implement in their curriculum (called 'the pilots'). Teachers took part with one whole class of students (ranging from 6 to 22), meaning that the whole class enrolled in the same assessment. Students had no choice in this, and teachers did not select specific students as participants. In total 453 students (206 boys, 247 girls; Mean Age = 14, 2 years) and 76 teachers took part. All students were at the end of their third year of study (the total study programme lasts four years) and had made a choice for one of the five occupational fields of green education. They conducted the new assessment in the field of their chosen track. A total of 68 students, ranging from 1 to 5 per pilot, voluntarily participated in an evaluation interview after conducting the assessment. All teachers ($n = 76$) participated in a structured intake discussion before assessment implementation and a group evaluation interview afterwards.

Every pilot was supported and coached by (1) a university researcher, the first author, with expertise in the field of assessment and thorough understanding of the new competence-based assessments in both VET and pre-VET, and (2) an 'expert teacher' from another green pre-VET institution with expertise in the field of the chosen assessment (e.g. floristry, animal and care). These 'expert teachers' participated in developing the new assessments within their field of expertise (see Part I). As such, they had a thorough understanding of both the subject area as well as the assessment principles underlying the new assessments. From now on, these 'expert teachers', 24 in number, are called 'coaches'. They were involved for the purpose of helping their colleague teachers, from another institution but teaching the same subjects, in thinking about how to implement the assessment as intended in their own context. The coaches participated in every step of the assessment implementation. The two project leaders of the Green Assessment project functioned as (1) sounding board for the researcher and coaches with whom they had monthly meetings and (2) organisers of national meetings for the green pre-VET institutions to share experiences on the new competence-based assessments. As such they inspired the pilot teacher teams.

Instruments

Structured intake discussion

Every pilot started with a structured intake discussion between the teacher team, researcher and coach. This served two goals (see also Gulikers, Baartman, and Biemans 2010): (1) systematically discussing the characteristics and steps of the new assessment and their underlying theoretical rationales for the purpose of creating a shared understanding of these aspects and how they relate to assessment quality (e.g. the importance of authenticity and combining assessment methods; why an observation scheme is put in place, etc.,). This was important as these new assessments were quite different from what teachers and managers were used to, and (2) building a shared understanding of the actual implementation of the chosen assessment in their own context to ensure that the assessments are implemented as intended as much as possible as this is crucial for the quality of the implemented assessments.

A structured open-ended questionnaire was developed explicating the characteristics and underlying rationales of this specific new assessment (see also Gulikers, Baartman, and Biemans 2010). For example, it asked ‘the assessment describes an assessment procedure for observing and evaluating students. Why do you think these procedures are like this? And how do you see this happening in your own context?’ The researcher guided the discussion and together with the coach fuelled the discussion with open and stimulating questions to let the team actively think about the what, why and how of the new assessment.

The structured intake discussions were typed out along the questions of the intake questionnaire and sent to the participating teacher teams for a member check. This provided comparable data on how the 34 pilots interpreted the assessment design characteristics, showing the degree to which teachers understood the intended ideas behind the new assessments and their expectations with respect to challenges and effects on teaching and learning.

Semi-structured group evaluation interviews

After every assessment two semi-structured interviews were held: one with a selection of students and one with the teacher team. Additionally, at the end of all pilots, one interview with the coaches was conducted. They focused on (1) the experienced effects of the assessment on students and teachers/assessors, (2) the assessment characteristics that led to these experienced effects; (3) what should be changed to develop an aligned, competence-based teaching-learning-assessment process.

All evaluation interviews were conducted by the same researcher (first author) and typed out. The teacher and coach interviews were sent to teachers and coaches, respectively to commend on and approve.

Open ended evaluation questions

The students participating in the evaluation interview individually answered four open questions on paper at the start of the interview. These questions were: what did you like, what was different than normal assessments/tests, what did you learn, and what did you not like.

Analysis

Three raters, the researcher and the two project leaders, independently analysed the data described above. The three individual analyses were compared and discussed in two intense reflection meetings to identify trends per stakeholder group and develop a collaborative line of argument for answering the research questions. The individual analyses of the intake reports were input for the first reflection meeting; individual analyses of the interview and questionnaire data were input for the second reflection meeting. Data analysis followed first a within case (per pilot) and then a cross-case (across pilots) analysis (Miles and Huberman 1994). The cross-case analysis focused on identifying trends, across pilots for the three stakeholder group separately (i.e. student, teachers and coaches). The same analysis questions and procedures were used for both the intake reports as well as the interview and questionnaire data, with the difference that the intake reports addressed the *expected* effects (as the assessment were not implemented yet) and only provided information from teachers (not students and coaches). The trends identified in both reflection meetings were combined to develop a collaborative argumentation for answering the research question per stakeholder group. Quotes from student, teacher, or coaches answers were used to support identified trends and line of argument when insightful.

A third intense reflection meeting was held between the researcher, project leaders and the coaches; the latter as representatives of the teachers in this action research. The developed argumentation for answering the research questions was elaboratively discussed for the purpose of validation (i.e. did the coaches recognise and agree with the arguments based on their own experiences of the pilots?), and adjusted or nuanced if the majority of the coaches deemed this necessary.

Results

Stakeholders' evaluation of the implemented assessment

This section describes the evaluations of the three stakeholder groups separately. The discussion will combine the evaluations in main trends about the flywheel function of the new assessment for changing teaching and learning

Students

First, students were univocally enthusiastic about the *motivational and formative effect* of assessments: 'I can really show what I can do. That is really motivating' or 'I tried several times to properly hoard the plants, but I did not seem to get it. I now know what to practice better the next time' (both gardening students). Students said that they actually did not specifically learn *for* the assessment but learned a lot *from* the assessment: 'I experienced for myself that I really developed the last couple of weeks even though I did not really learn for the assessment' or 'unconsciously you learn to apply everything you have learned in a practical situation' (floristry students). Second, students explicitly linked the *authentic* task of the assessment, which was much more meaningful and realistic than in previous assessments, to this motivational and formative effect and their preparation for VET.

the assessment really showed me what being a gardener means and I really liked helping with constructing this garden for a real client. Actually, I now seriously consider to follow up my pre-VET studies with a VET study in gardening. (gardening student)

On the other hand, they explained that their experienced authenticity, and resulting motivation, depended on: (1) if there was a real or a 'fake' client involved (e.g. a teacher playing the role of client), and (2) if students had to make a real product for this client that was actually given to the client.

We had to make a flower bouquet for the elderly people in the geriatric home behind our school. However, when I returned to class one week after the assessment, all our bouquets were left to rot there. Why weren't they picked up or delivered to the geriatric home? It was as if we did it all for nothing. (floristry student)

Third, with respect to the *alignment to education* many students reported that their preparatory education did not (yet) properly prepare them for this new assessment in various ways: (1) in class they never worked on or discussed competencies, while the assessment assessed competencies; (2) the assessment required student to make independent decisions and deal with complex situations themselves, whereas in class they were used to being told what to do step-by-step; (3) some assessment activities, like dealing with customers, were not practiced much during classes, and (4) students were not used to reflect on their performance in a way that was expected in the assessment interview. They expected the assessment interview to be more like an oral exam focusing on content-related theoretical questions.

Teachers

First, teachers had low expectations of their students beforehand and they were enormously surprised by what their students were able to do in the assessments. For instance, in the intake discussions, many teachers appeared sceptical about their students' ability to plan, deal with the complexity of the authentic assessment task perform these tasks independently and their ability to reflect on their performance in the assessment interview: 'planning and reflecting with pre-VET students??? I think that is too much to ask' (gardening teachers). 'these students

need a lot of guidance. You need to take them by the hand. I wonder if they can deal with an assessment task that requires them to do multiple things during four hours' (food teacher). After the assessments were conducted, teachers were specifically surprised by the degree of independence students showed, their creativity in solving problems, their capacity to reflect on their own performance and the enthusiasm exposed.

I am really surprised with what these kids are able to tell me in the assessment interviews. When you ask the right questions they seem to be much more conscious of their own performance than I ever thought they would be. (gardening teachers)

Second, all teachers experienced the assessment tasks to be very *authentic* and *meaningful* for students' future, both with respect to career as well as further VET education, motivating students to perform at their best.

The most important innovation of this new assessment, I think, is that it requires students to work with real clients on real occupational tasks. I think that makes the assessment much more motivating for our practically-oriented students (Floristry teachers. Intake discussion)

...one of the girls in my gardening class told me that this assessment showed her how much fun the gardening job can be. As a result of that she seriously considered going to VET in Gardening, while before she always thought to proceed to VET in Care and Welfare. (Gardening teacher)

Third, the most evident challenge reported by all teachers (100% of the teachers expressed this evaluation) related to their way of teaching: 'if I want to properly use these assessments in our programme, I really need to change how I teach and approach my student'. Teachers explained that they were used to take students by the hand in a step-by-step manner, while these assessments required students to oversee multiple smaller tasks and take initiative in planning and independently dealing with a complex task. Both teachers and students were not used to this way of working. Teachers' evaluation saying 'I need to learn to keep my mouth shut!' (100% of the teachers reported an evaluation like this) was exemplary of the change process required in teachers' pedagogy; teachers realised they needed to learn to not immediately help and interrupt students whenever they seem in doubt or have a question.

sometimes students were inclined to ask my help (i.e. the teacher being the assessor) when they were not sure what to do during the assessment. I tried not to do that even though that was hard for me. I was surprised that most students just solved the issue themselves, without complaining (animal care teacher)

Also content-wise teachers realised the need to change *what* they taught students. For example, 'I did not cover the aspect of sustainability in my class properly' (food teacher) or 'I never practiced with having reflective interviews with students' (several teachers) or 'I did not pay careful and explicit attention to dealing with customers' (several teachers).

Fourth, teachers did not yet have thorough understanding of what competencies are, how they can be observed in student performance or discussed in

assessment interviews. Teachers raised questions like: 'How do you observe generic competencies in a certain context?' and 'What do I want to hear students tell me about these competencies in the assessment interview?'. Teachers were used to testing theoretical knowledge and ticking of practical skill performance (i.e. milking cows, following hygiene procedures) but they felt not yet confident in observing and talking about competencies. Even though the assessment contained an observation scheme and interview questions that operationalised the four to be assessed competencies teachers felt insufficiently skilled for assessing students' competencies.

Coaches

The themes identified in the evaluations of the coaches were largely comparable to the teacher evaluations with respect to the motivational effect of the authentic assessment task and the misalignment to both the teaching content and pedagogy. Additionally, the coaches valued the *assessment form*, that is a combination of performance of a situated task combined with an assessment interview to assess student' generic competencies, as an important preparation for students' future in competence-based VET. Green VET institutions all used summative assessments of this same form, in VET often conducted in students' internships.

The coaches discussed the extent to which teachers *implemented the assessment as intended*, referring mainly to teachers' difficulties in observing and discussing students' competencies during the assessments. Coaches argued that this became evident in the assessment interviews, both in the kind of questions assessors asked and the way they asked them. Questions often focused on theoretical knowledge, instead of asking student to explain something about their performance of certain competencies (e.g. 'can you tell me what kind of flowers fit in a bouquet together with roses' instead of 'you advised the customer to use hot water for this bouquet, while you did not explain why that was so important. Can you explain why you handled this situation this way?'). With respect to the way assessors asked questions, coaches argued that assessors mostly asked closed, instead of open questions. And that they did not elaborate on students' answers. For example

Teacher: Do you think you satisfied the customer well enough?

Students: 'yes'

After which the conversation stopped. Using the assessment interview to provide additional information on students' competencies would require a follow up question like:

Teacher: 'what did you do to find out the customers wishes' and/or 'how did you address these in your performance?'

The coaches stressed the importance of more professionalisation of their colleague teachers in how to properly implement the new competence-based assessment method as intended, certainly with respect to observing and discussing competencies.

Conclusions and discussion

This article examines a competence-based assessment innovation project in pre-vocational secondary education (pre-VET), preparing for competence-based VET. It substantiates the design of a new authentic competence-based performance assessment and examines how implementing such a new assessment can be a flying wheel for changing teaching and learning towards more competence-based education. Four conclusions can be drawn relating to required changes in teaching content and pedagogy or professional development necessary for developing aligned competence-based practices in the future.

First, collaboratively discussing the theoretical rationales and characteristics of the new assessment before implementing the assessment, is critical for building a shared understanding of what this new assessment looks like, how it is different from what the teachers are used to and how they want to implement it within their own context. This is important for increasing the extent to which the assessment is implemented as intended, which is prerequisite for changing towards competence-based practices. In line with previous experiences (Gulikers, Baartman, and Biemans 2010; Gulikers et al. 2013) the intake discussions in which teacher teams were supported by experts and guided by a structured questionnaire was helpful for reaching these goals. This finding agrees with the raising number of action research approaches in assessment innovation projects (e.g. Gulikers et al. 2013; Torrance and Pryor 2001)

Second, changing towards competence-based assessment requires careful attention to developing understanding of what competencies are and how they can be assessed. Teachers in our study were still unsure about this during the assessments, let alone that they changed their teaching from being knowledge-oriented towards more competency-oriented. The intake discussion and all the guidelines provided in the assessment (e.g. the observation scheme, the contextualisation of the four competencies, the guiding questions for the assessment interview) were helpful, but not enough for teachers to develop specific images of what competencies actually mean in their own teaching practice (de Bruijn and Leeman 2011). This again challenges the extent to which the assessment is implemented as intended. Helping teachers building up practical knowledge of competencies and how to assess them should be at the heart of future interventions and professional development. For instance, using moderation and calibration techniques to learn how to qualify student performances should be used in this respect more often (Grainger, Purnell, and Zipf 2008; Sturgis 2014).

Third, teachers and coaches were surprised by what their pre-VET students were able to show in their performance and assessment interviews, while students reported that the assessment allowed them to show what they can do or that the assessment made them aware of everything they unconsciously learned. Two conclusions can be drawn from this challenging finding. First, this seems to suggest that the new assessments are actually measuring something different

than the formal curriculum, which can be seen as the proof of the formative effect of the assessment itself. Either, the assessed competencies are implicitly and/or unconsciously developed during the curriculum and made explicitly visible by the assessment, or knowing that the new competence-based assessment is coming up, had a pre-assessment formative effect (Gielen, Dochy, and Dierick 2003) on teachers' teaching and students' learning. Second, it stresses the low expectations that teachers have of their pre-VET students. We argue that both teachers and students would profit a lot from having higher expectations of students. Hattie's well-known meta-analysis (2009) as well as monitoring studies from the Dutch Inspectorate of Education (2011) showed that schools where teachers have high and ambitious (but attainable) expectations of their students, have higher learning outcomes. Certainly from the perspective of smoothing the transition from pre-VET to VET and developing more continuous learning pathways it would be wise to look at the expected learning outcomes and processes of mainstream VET. These would facilitate developing these high, but attainable learning outcomes for pre-VET students in the final stages of their education.

Fourth, in line with the low expectations argument, 'I have to learn to keep my mouth shut' (quote of many teachers) is indicative of the most prominent pedagogical challenge for teachers. Teachers tended to intervene during students' performances whenever they saw students doubting about what to do next. In response, students expect to be taken by the hand and immediately ask questions whenever in doubt. This was clearly not what was intended by the assessment designers; this challenges a fair assessment of students' own and independent performance of the whole task. This issue, showing the current student-teacher interaction pattern, is at the heart of the misalignment of the new assessment with current 'traditional' education that all stakeholders report. Pre-VET education is pre-structured and characterised by teachers guiding students step-by-step and breaking up difficult (whole) tasks into small pieces. In competence-based education, students should be challenged more to perform complex, vocational problems and guide their own learning process without being this dependent on their teachers (Sturing et al. 2011). Thus, the way students and teachers interact in their regular educational practice needs to be addressed for a proper implementation of these assessments as an integral part of the curriculum (Misbah et al. 2015) as well as for preparing students for independent performance of occupational tasks expected in VET (EQF level 4; see described in the 'research context').

In sum, all stakeholders became very engaged in the new assessments. As such, it already is a successful innovation, even though the assessment is not fully implemented as intended yet.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Achtenhagen, F. 2012. "The Curriculum-Instruction-Assessment Triad." *Empirical Research in Vocational Education and Training* 4 (1): 5–25.
- Achtenhagen, F., and E. Winther. 2014. "Workplace-based Competence Measurement: Developing Innovative Assessment Systems for Tomorrow's VET Programmes." *Journal of Vocational Education & Training* 66 (3): 281–295.
- Baartman, L. K. J., Th J. Bastiaens, P. A. Kirschner, and C. P. M. van der Vleuten. 2006. "The Wheel of Competency Assessment: Presenting Quality Criteria for Competency Assessment Programmes." *Studies in Educational Evaluation* 32: 153–170.
- Baartman, L., and J. Gulikers. 2017. "Assessment in Dutch Vocational Education: Overview and Tensions of the past 15 Years." In *Enhancing Teaching and Learning in the Dutch Vocational Education System: Reforms Enacted*, edited by E. de Bruijn, S. Billett, and J. Onstenk, 245–266. Dordrecht: Springer.
- Biggs, J. 1996. "Enhancing Teaching through Constructive Alignment." *Higher Education* 32: 347–364.
- de Bruijn, E., and Y. Leeman. 2011. "Authentic and Self-directed Learning in Vocational Education: Challenges to Vocational Educators." *Teaching and Teacher Education* 27 (4): 694–702.
- De Jonge, L. P. J. W. M., A. A. Timmerman, M. J. B. Govaerts, J. W. M. Muris, A. M. M. Muijtjens, A. W. M. Kramer, and C. P. M. van der Vleuten. 2017. "Stakeholder Perspectives on Workplace-based Performance Assessment: Towards a Better Understanding of Assessor Behaviour." *Advances in Health Sciences Education*. doi:10.1007/s10459-017-9760-7.
- Dietzen, A., and R. Nicolaus. 2016. "Measuring and Modelling Occupational Competences: Practical Implications." Symposium presented at the Competence 2016 Conference. Wageningen, The Netherlands.
- Gielen, S., F. Dochy, and S. Dierick. 2003. "Evaluating the Consequential Validity of New Modes of Assessment: The Influence of Assessment on Learning, including the Pre-, Post-, and True Assessment Effects." In *Optimising New Modes of Assessment: In Search of Quality and Standards*, edited by M. Segers, F. Dochy, and E. Cascallar, 37–54. Dordrecht: Kluwer Academic Publishers.
- Grainger, P., K. Purnell, and R. Zipf. 2008. "Judging Quality through Substantive Conversations between Markers." *Assessment & Evaluation in Higher Education* 33 (2): 133–142.
- Gulikers, J. T. M., L. K. J. Baartman, and H. J. A. Biemans. 2010. "Facilitating Evaluations of Innovative, Competence-Based Assessments: Creating Understanding and Involving Multiple Stakeholders." *Evaluation and Program Planning* 33 (2): 120–127.
- Gulikers, J. T. M., T. J. Bastiaens, and P. A. Kirschner. 2004. "A Five-dimensional Framework for Authentic Assessment." *Educational Technology Research and Development* 52 (3): 67–86.
- Gulikers, J. T. M., Th J. Bastiaens, and P. A. Kirschner. 2006. "Authentic Assessment, Student and Teacher Perceptions: The Practical Value of the Five-dimensional Framework." *Journal of Vocational Education and Training* 58: 337–357.
- Gulikers, J., H. Biemans, and M. Mulder. 2009. "Developer, Teacher, Student and Employer Evaluations of Competence-based Assessment Quality." *Studies in Educational Evaluation* 35 (2-3): 110–119.
- Gulikers, J. T. M., H. J. A. Biemans, R. Wesselink, and M. van der Wel. 2013. "Aligning Formative and Summative Assessments: A Collaborative Action Research Challenging Teacher Conceptions." *Studies in Educational Evaluation* 39 (2): 116–124.
- Harsch, C., and G. Martin. 2012. "Comparing Holistic and Analytic Scoring Methods: Issues of Validity and Reliability." *Assessment in Education: Principles, Policy & Practice* 20 (3): 281–307.
- Hattie, J. 2009. *Visible Learning. A Synthesis of Over 800 Meta-analyses Relating to Achievement*. New York: Routledge.

- Hawkins, R. E., C. M. Welcher, E. S. Holmboe, L. M. Kirk, J. J. Norcini, K. B. Simons, and S. E. Skochelak. 2015. "Implementation of Competency-based Medical Education: Are We Addressing the Concerns and Challenges?" *Medical Education* 49 (11): 1086–1102. doi:10.1111/medu.12831.
- Inspectorate of Education. 2011. *Opbrengsten, maak er meer van!* [Performances, Make the Most out of Them]. The Hague: Ministry of Education.
- Knight, P. 2000. "The Value of a Programme-wide Approach to Assessment." *Assessment and Evaluation in Higher Education* 25 (3): 237–251.
- McMullan, M., R. Endacott, M. A. Gray, M. Jasper, C. M. I. Miller, J. Scholes, and C. Webb. 2003. "Portfolio and Assessment of Competence: A Review of the Literature." *Journal of Advanced Nursing* 41 (3): 283–294.
- Miles, M. B., and A. M. Huberman. 1994. *Qualitative Data Analysis: An Expanded Sourcebook*. Thousand Oaks, CA: Sage Publications.
- Miller, G. 1990. "The Assessment of Clinical Skills/Competence-/Performance." *Academic Medicine* 65 (9): S63–7.
- Misbah, Z., J. Gulikers, R. Maulana, and M. Mulder. 2015. "Teacher Interpersonal Behaviour and Student Motivation in Competence-based Vocational Education: Evidence from Indonesia." *Teaching and Teacher Education* 50: 79–89. doi:10.1016/j.tate.2015.04.007.
- Mulder, M. 2014. "Conceptions of Professional Competence." In *International Handbook of Research in Professional and Practice-based Learning*, edited by S. Billett, C. Harteis, and H. Gruber, 107–137. Dordrecht: Springer.
- Popham, W. J. 2001. "Teaching to the Test?" *Educational Leadership* 58 (6): 16–21.
- Sturgis, C. 2014. "Progress and Proficiency: Redesigning Grading for Competency Education." Competency Works Issue Brief. International Association for K-12 Online Learning. Accessed July 3, 2017. <https://eric.ed.gov/?id=ED561319>
- Sturing, L., H. Biemans, M. Mulder, and E. de Bruijn. 2011. "The Nature of Study Programmes in Vocational Education: Evaluation of the Model for Comprehensive Competence-based Vocational Education in the Netherlands." *Vocations and Learning* 4 (3): 191–210. doi:10.1007/s12186-011-9059-4.
- Torrance, H., and J. Pryor. 2001. "Developing Formative Assessment in the Classroom: Using Action Research to Explore and Modify Theory." *British Educational Research Journal* 27 (5): 615–631.
- Van Merriënboer, J. J. 1997. *Training Complex Cognitive Skills: A Four-Component Instructional Design Model for Technical Training*. New Jersey: Educational Technology.
- Weigel, T., M. Mulder, and K. Collins. 2007. "The Concept of Competence in the Development of Vocational Education and Training in Selected EU Member States." *Journal of Vocational Education and Training* 59 (1): 51–64.
- Wesselink, R., H. J. A. Biemans, J. Gulikers, and M. Mulder. 2017. "Models and Principles for Designing Competence-Based Curricula, Teaching, Learning and Assessment." In *Competence-based Vocational and Professional Education. Bridging the Worlds of Work and Education*, edited by M. Mulder, 533–553. Cham: Springer.
- Young, M. F. D. 2009. "National Qualifications Frameworks: An Analytical Overview." In *International Handbook of Education for the Changing World of Work*, edited by R. Maclean and D. Wilson, 2867–2880. Bonn: Springer.