



18/10/2023

# Assessment in O&O Education

Supporting O&O Students in their Development

Irene Wols – s1912623

Master Thesis ECB-O&O

DPM2051

Dr. Ir. Robert-Jan Den Haan

Prof. Dr. Ir. Masha van der Voort

Jony Heerink MSc

## Abstract

Providing high quality assessment in the secondary school subject of Onderzoek & Ontwerp (O&O; Research and Design) is challenging for O&O-teachers according to previous research. Despite dedicating substantial time to assessment, teachers express doubts about the educational effectiveness of their efforts. Additionally, there is a lack of consensus among teachers regarding what should be assessed and how to assess it.

This report, therefore, aims to design a way to support teachers in their assessment in O&O. To gain deeper insight into current assessments practices in O&O, obstacles faced in assessment in O&O, quality criteria for assessment and potential solutions, interviews with O&O-teachers, O&O-students and university staff were conducted. Afterwards a focus group session with O&O-teachers was organised to validate the results of the interviews and identify potential solutions.

Inductive thematic analysis of the responses to the interviews and the focus group resulted in three distinct purposes for assessment in O&O: (1) quality assurance, (2) motivation and (3) student development. Further analysis of the results indicates that the development of a solution that utilizes assessments to support student development was most relevant. Reframing the research question to fit this direction, resulted in the design of six concept solutions. These concepts were then evaluated in two steps by two different teachers, resulting in one final design. This 'Challenge Track' uses the principle of 'creating space for growth,' assessing the time and effort students invest in their development rather than the development itself.

## Table of Contents

_Toc147991041Introduction .....	3
Theoretical Framework.....	5
Assessment in education .....	5
O&O education .....	10
Study objectives .....	11
Methods.....	13
Ethical considerations .....	13
Participants .....	14
Procedure.....	15
Analysis .....	21
Results interviews and focus group .....	22
Quality Assurance .....	22
Motivation.....	28
Student Development.....	31
Tensions .....	38
Reframing.....	41
Assessment purposes.....	41
Updated Model.....	44
Reframing.....	45
Design.....	47
Initial concepts.....	47
First evaluation .....	51
Selected concepts .....	56
Second evaluation.....	58
Final design .....	62
Discussion & Conclusion .....	65
Limitations .....	66
Recommendations .....	67
References .....	69
Appendix .....	72
Appendix A: Overview of Quality Criteria .....	72
Appendix B: Overview of Obstacles.....	74
Appendix C: Overview of Solutions.....	74

## Introduction

Assessment plays an important role in making informed decisions about the future of students and educational courses. To make these decisions, assessment by and large entails checking what and how much students learn (AERA et al., 2014). Learning itself is however not a phenomenon that can be empirically observed, because, as we all intuit, it is something that happens within our minds. Therefore, the occurrence of learning in the mind of a student can only be inferred through changes in observable behaviour after learning has taken place (Biesta, 2022). This inference from observable behaviour is the essence of assessment (Baird et al., 2017).

Because important decisions about the future of students and educational courses can depend on assessments, it is important to develop these assessments with care and attention. Therefore, guidelines to develop proper assessment strategies have been systematically formulated in traditional educational settings (Hodges, 2013); i.e., courses that follow a tightly structured educational style, such as lectures in which teachers transmit knowledge to the students and the students are mostly passive listeners. The most prominently used guidelines in the United States can be found in the 'Standards for Educational and Psychological Testing' (The Standards) (AERA et al., 2014). While these forms of guidelines work particularly well for the mentioned educational settings, they are seen as less suitable for some novel educational courses, such as challenged-based learning (CBL) courses (Schuwirth & Van der Vleuten, 2012).

One of the main differences between CBL courses and traditional courses is the way the learning goals are conceptualized (Gallagher & Savage, 2023). These learning goals do not only focus on the attainment of disciplinary knowledge, but also on the personal and professional development of the students. The assessment of these types of learning goals has been considered complex and subjective and to date no satisfactory guidelines have been developed (Boor et al., 2021; Chan & Luk, 2022; Hughes & Barrie, 2010).

An example of a CBL course that requires new assessment guidelines is Onderzoek en Ontwerp (O&O; Dutch for Research and Design), a secondary school course taught at some high schools in the Netherlands. In O&O, students work in small groups on authentic, STEM-related (Science, Technology, Engineering, Mathematics) challenges. The learning goals of O&O focus on the development of skills or competencies (Schalk & Bruning, 2014; Stichting Technasium, 2021). As can be anticipated based on these learning goals, previous research shows that teachers struggle with assessment in O&O (Schalk & Bruning, 2014; Visser, 2023; Wols, 2022). This research points out various issues, including:

- Teachers are uncertain about how to use assessment to support the development of students (Wols, 2022).
- Considerable time is spent on assessment, yet teachers remain doubtful about the value of this investment (Wols, 2022).
- There is a lack of consensus among different schools and sometimes even within the same schools regarding preferred assessment strategies (Visser, 2023).
- The substantial involvement of teachers in O&O-projects can influence the quality of the students' work (Schalk & Bruning, 2014).

These issues have a direct impact on students, as the assessment and grading of students in O&O plays a role in their placement and graduation.

This report aims to contribute to tackling the problems teachers face with assessment in O&O through the design of a tool. To achieve this goal, methods from Human-Centred Design (HCD) are adopted to ensure that the developed tool aligns with the needs and preferences of the users. In this project, O&O teachers are considered the main users, since identified issues are primarily teacher related. Therefore, the design process is guided by the following research question:

*What tool can be designed to effectively support O&O teachers in providing high-quality assessments in O&O?*

To answer this research question, the report first presents an examination of existing literature on assessment and O&O education. Afterwards, the report outlines the chosen methodology, which is followed by an overview of the most important results from interviews and a focus group. Insights from these results prompted a reframing of the initial research question. Building upon this, the report presents a design process in which concept designs are developed and evaluated, resulting in a final design. Finally, the report is concluded with a discussion of the final design and recommendations for next steps.

## Theoretical Framework

In this section, literature is examined to provide a background on the following topics: (1) assessment of learning in traditional education and CBL; and (2) O&O education in the Netherlands. At the end of the chapter, sub-questions to the research question are developed.

### Assessment in education

This section begins by establishing a definition of assessment, drawing from existing literature. This definition serves as a foundation for situating assessment within the broader context of education, leading to the construction of a conceptual model. Subsequently, the section delves into traditional quality criteria for assessment. Finally, two examples of novel assessment strategies are provided.

### Definition of Assessment

As described in the Introduction, assessment at its core is an attempt to find out what a student has learned or how competent a student is (Baird et al., 2017; Wiliam, 2011). Although sharing this core, assessment has been defined and is used in slightly different ways in different papers. For the sake of clarity, in this report assessment is defined as the following:

Assessment: The process of first *checking* and then making *decisions* based on the extent to which students have mastered certain *constructs*.

This definition has three important concepts that require further elaboration: (1) *constructs*, (2) *checking* and (3) *decisions*.

A *construct* is a concept or characteristic that an assessment intends to measure (AERA et al., 2014). Similar to learning, a construct itself is not a phenomenon that can be empirically observed; however, its existence can be inferred from observable behaviour (Baird et al., 2017; Biesta, 2022). For example, if a test intends to measure the reading comprehension of students, the construct would be 'reading comprehension'. Reading comprehension itself cannot be observed; however, it can be inferred from a students' answers to questions about a text.

*Checking* involves the systematic collection of evidence to make an inference about a student's mastery of the construct (Schuwirth & Van der Vleuten, 2012; Wiliam, 2017a). This evidence ranges from observations of what students do and make, to answers to test questions (Wiliam, 2017a). The quality of the evidence and the inferences is described in terms of validity, reliability and fairness, which will be discussed in detail later.

*Decisions* refers to the consequences of the results of an assessment, such as passing or failing a course (Baird et al., 2017). A well-structured assessment intends to make specific decisions. This intention is considered the 'purpose' of an assessment (AERA et al., 2014; Newton, 2007). Examples of these purposes include qualification, selection and institution monitoring (Newton, 2007).

### Situating assessment in Education

Assessment is inherently embedded in the larger system of education, which is comprised of different elements. To create a coherent education system that supports high-level learning, the different elements of the educational curriculum need to be aligned and considered together (Biggs, 2003; Pellegrino & Chudowsky, 2003; van den Akker, 2003). Here once more, various authors have mentioned diverse lists of curricular elements. Two of these elements that are frequently mentioned in conjunction with assessment are (1) *goals* and (2) *learning activities* (Baird et al., 2017; Black & Wiliam, 2009).

The *goals* of an educational course represent the point on the horizon students are working towards (Biggs, 2003; Black & Wiliam, 2009). These are often formalised in intended learning outcomes or learning objectives. In a well-structured curriculum, assessment intends to measure to what extent students have reached these goals (Black & Wiliam, 2009). Consequently, goals effectively define the constructs to be measured (Wiliam, 2017b).

Nevertheless, the intention of educational courses is not just to measure the extent to which students have mastered a certain construct. These courses also intend to facilitate the improvement of students in the constructs (Baird et al., 2017). *Learning activities* refer to actions undertaken by teachers and students with the intention of improving the student’s mastery of the construct.

In the definition given above, it was described that assessment is the process of checking construct mastery. Through this, assessment design operationalises constructs. This is because assessments take the abstract idea of a construct, such as ‘mathematical ability’ and use this to develop concrete set of tasks, such as a specific set of mathematical problems for students to solve (Baird et al., 2017; Wiliam, 2017a).

Figure 1 depicts a developed conceptual model illustrating the assumed relationships between the goals, learning activities and assessment. Constructs are placed at the heart of this model, serving as the element that connects the goals, learning activities and assessment practices. This model furthermore shows that the different curricular elements need to be aligned with each other, as argued above. Decisions are made based on assessment outcomes. This model serves as a conceptual framework for the remainder of this research, enabling structured analysis of current and future practices in O&O.

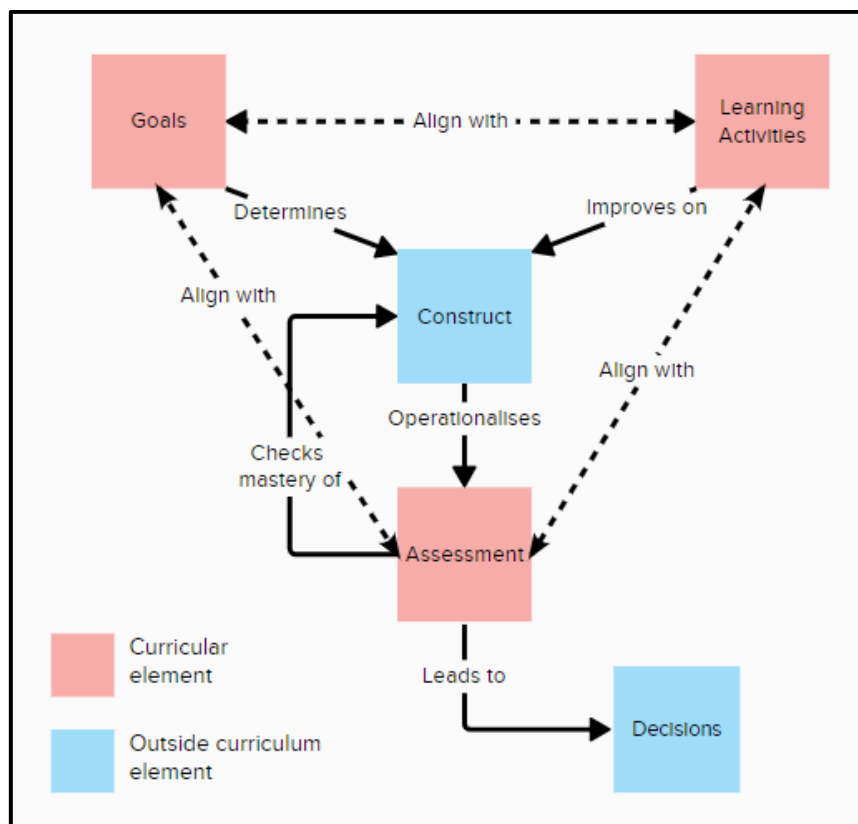


Figure 1: Model of situating assessment in educational courses. The curricular elements of assessment, goals and learning activities are aligned to create a coherent system. The constructs serve as a central connecting element between these curricular elements. The assessment results in decisions.

## Quality Criteria

To determine the quality of an assessment, the literature points towards several crucial factors. First the two factors emerging from the model in Figure 1 are discussed: (1) *construct clarity* and (2) *washback*. Following that, based on the descriptions in 'The Standards' three factors from the field of psychometrics are explored: (1) *validity*, (2) *reliability* and (3) *fairness* (AERA et al., 2014). In addition to these factors, three types of assessment norms are discussed. Lastly, some tensions are discussed.

### *Construct clarity*

Since constructs serve as a linking pin in the curriculum, an important first step when setting up assessments is defining the constructs to be measured. These constructs should be defined as precisely as possible, because well-defined constructs simplify assessment design (AERA et al., 2014; Wiliam, 2017a). On the other hand, failing to properly define constructs can give rise to assessment issues. These issues include disagreements about interpretations of assessment results and the inclusion of assessor bias (Wiliam, 2017a). In summary, it becomes difficult to perform a high quality assessment when it is unclear what is being assessed (Schuwirth & Van der Vleuten, 2012).

Clarity of constructs is furthermore important to ensure transparency for all stakeholders involved, including teachers, school principals, students, parents and educational inspectors (Broadfoot & Black, 2004). This clarity serves to establish a common understanding of what is being assessed and what is considered a sufficient result (Schalk & Bruning, 2014).

### *Washback*

In a well-designed course, students are enticed to engage in those learning activities necessary to achieve the learning goals (Biggs, 2003). This enticement can be done through the assessment practices, as these play a role in determining what students consider to be important (Biggs, 2003; Hughes & Barrie, 2010). This influence of assessment on learning activities is called washback (Green, 2007). Washback can have both positive and negative consequences for student learning, therefore assessments should be carefully designed to motivate those behaviours that lead to a better mastery of the construct instead of superficial learning to just improve assessment results (Baird et al., 2017).

### *Validity*

Validity refers to the degree to which an assessment effectively captures the intended construct. In other words, to what extent an assessment measures what it intends to measure. It is important to note that an assessment by itself cannot be either valid or invalid, as the validity depends on the purpose and intended interpretation of the assessment. Validation of an assessment is the process of checking whether there is sufficient evidence and theory available to support the interpretations of a test.

Problems with validity can have two main causes. The construct can be underrepresented or there can be construct-irrelevant variance. Construct underrepresentation refers to an assessment not capturing all important aspects of the construct. For instance, an assessment designed to measure French language ability that only focusses on reading skills, neglecting listening, speaking and writing abilities.

Construct-irrelevant variance refers to the outcome of an assessment being influenced by factors unrelated to the construct. An example would be when a test that intends to assess mathematical ability necessitates strong reading skills. Students with weak reading skills might fail the test regardless of their mathematical ability.



### *Reliability*

Reliability refers to the extent to which an assessment produces consistent and stable results. Simply put, it measures the degree to which an assessment, when administered a second time to the same group of students, yields consistent results. Variations in these results can be attributed to a range of factors, including:

- Timing of the assessment;
- Items used in the assessment;
- Examiners scoring the assessment.

In general, reliability relates to those elements that are assumed to not have any effect on the outcome of the assessment, but actually do.

Standardized tests have the advantage of high reliability, as they give the same or very similar test materials to all test takers in the same or very similar conditions with a very strict scoring procedure that can be applied consistently. This is why these tests are often used in examinations or other high-stake settings.

### *Fairness*

Fairness entails ensuring that all students get equal opportunity to demonstrate their ability in the targeted construct. This means that the test is accessible to all intended examinees irrespective of factors such as disabilities, race or gender. In practice, this may require the adaptation of tests to accommodate the needs of individual students. For instance, offering an auditory version of the test to visually impaired students.

### *Norms*

In assessment roughly three different types of norms can be distinguished: (1) Growth-based, (2) Normative and (3) Standards-based (AERA et al., 2014). Growth-based norms assess students on their performance compared to their previous performance, judging the progress they made. Normative norms compare students to their peers, indicating their position in the population of students participating in the assessment. Lastly, standards-based norms assess students based on some standards. These standards can be established by, for example, a committee of experts in the area to be tested.

### *Tensions*

Because standardized tests lead to high reliability, the late twentieth century saw a push for standardization in assessment (Hodges, 2013). This caused a tendency to rule out any form of subjectivity, as subjectivity was regarded as biased and unfair (Van der Vleuten et al., 1991). This tendency did however cause problems as well, as high levels of reliability do not necessarily result in high levels of validity (Hodges, 2013). Some aspects of a construct can be more easily captured in a standardized way than other aspects, which leads to a tendency for construct underrepresentation. When testing French language capability, it is easier to standardize the ability to answer multiple choice questions after reading a text, than the ability to have a conversation in French. Ironically, quality assurance procedures often exacerbate this problem (Biggs, 2003).

## Assessment strategies

Besides traditional forms of assessment, such as standardized tests, many other forms of assessment strategies have been developed. Two approaches that are particularly relevant to this report are discussed below: (1) the *jury model* and (2) *formative assessment*.

### *Jury model*

One of the quality criteria described before, is the clarity of construct that are intended to be assessed. However, several authors argue that constructs related to personal and professional development, such as teamwork or creativity, are often difficult to define specifically and concretely (Gibb, 2014; Hughes & Barrie, 2010). To avoid problems with assessment design, there has been a tendency to deconstruct these competencies into smaller and smaller sub-competencies that allow for precise construct descriptions (Hodges, 2013). The combination of the assessments of these sub-competencies intends to represent the student's competency in the initial construct (Biggs, 2003; Hodges, 2013).

However, both Biggs (2003) and Hodge (2013) have argued such an approach lacks the holistic perspective required for a comprehensive assessment of a student's mastery of the construct. Instead, they argue that in the case of hard-to-define fuzzy constructs, it is important to provide a holistic judgement of the student and to leave in a 'human component' (Biggs, 2003; Hodges, 2013). Or as Hodge (2013) puts it 'capturing the wisdom in holistic supervisor judgements'.

Hodge (2013) does however acknowledge that the judgement of just one supervisor can be biased and therefore unreliable. He instead argues for using judgements from multiple people in a panel of judges, the so-called 'jury model'. This model allows for the inclusion of subjectivity on the one hand, but through averaging the judgement of the panel members still come to reliable conclusions (Van der Vleuten et al., 1991). In the case of creativity, it has indeed been shown that this method results in reliable conclusions (Long & Wang, 2022).

### *Formative assessment*

Formative assessment is a wide spread term used both in academics and in practice, sometimes interchangeably with Assessment for learning (Sluijsmans, 2020; Wiliam, 2017a). This wide spread use has also caused the presence of many misconceptions and definitional unclarity (Sluijsmans, 2020), which Wiliam (2017b) has described this as a 'profound lack of agreement about what either of these terms mean'. For sake of clarity, in this report the definition as provided by Black & Wiliam (2009) will be used, which states in short that formative assessment uses evidence to make decisions about instruction aiming to better suit the educational needs of students.

Feedback is an important component of formative assessment, providing information to the student regarding an aspect of their performance or understanding (Hattie & Timperley, 2007). Specifically, feedback points towards the gap between what a student can do or understand and what is intended for a student to do or understand (Hattie & Timperley, 2007). Because of this, it is again important to clearly define the learning goals.

Formative assessment is the opposite of summative assessment, which is the type of assessment that takes place after the completion of a course for purposes such as certification, selection or admission purposes (Folmer et al., 2017; Newton, 2007).

The summative outcomes of the evaluation serve a communicative role between educational institutions and the larger world, as described by Broadfoot & Black (2004). Evaluative outcomes are in this case often publicly accepted as a code for quality: A diploma indicates a certain level of

competence. It is therefore important for summative assessment that there is consistency in the interpretations of multiple assessors, safeguarding reliability (Wiliam, 2017a).

For formative assessment this is however less important, as the focus lies on student learning (Wiliam, 2017a). Different interpretations from different assessors can point to different problems students might be experiencing and paint a holistic picture of what students and teachers can do to improve educational outcomes (Wiliam, 2017a).

### O&O education

In 2004 the Technasium foundation (Stichting Technasium) was set-up among other reasons to combat the shortage of STEM graduates in the Netherlands. (De Vijlder et al., 2014). To this end, the subject O&O was implemented in more than 100 secondary schools throughout the Netherlands (Stichting Technasium, n.d.). Students at these schools can usually pick O&O as an elective course. Below an overview of the general characteristics of this subject will be given using the elements of the model in Figure 1 (Stichting Technasium, 2021).

### Learning activities

Students choosing O&O, usually spent around five class hours per week on their projects. A client from industry is usually involved in these projects, providing an authentic research or design problem for students to solve (Blume-Bos et al., 2020). After a general introduction to the project and, if possible, an excursion to the company, students work on their projects in teams of three or four students (Prins et al., 2011). To make ensure that the projects are as authentic as possible, students should be working on unresolved questions, meaning that the projects are constantly changing (Stichting Technasium, 2021). As teachers cannot be content experts in each of these projects, they instead take on a coaching role to support the students during their projects (Lechner, 2012). Because the content knowledge does not come from teachers, the client and sometimes external experts are available to offer content support to the students and answer any questions they have about the content of the project (Prins et al., 2011; Stichting Technasium, 2021). The project usually ends with students presenting their results to the client (Lechner, 2012).

Technasium schools offer O&O from the first to the final year of secondary school, for a total of either five or six years. In the first three years of secondary school (onderbouw), students get three or four projects per year, each taking about eight weeks to complete (Lechner, 2012). In the final two or three years (bovenbouw) students work on half a year projects, finishing up with a full year project called 'meesterproef' in their final year (Prins et al., 2011).

### Goals

O&O has the overarching goals of preparing students for beta-technical studies and jobs (Schalk & Bruning, 2014; Stichting Technasium, 2021). Interesting O&O does not have learning goals that are linked to content knowledge, but instead the learning goals focus on skills or competencies students should develop themselves in (Schalk & Bruning, 2014; Stichting Technasium, 2021). These competencies include for example: team work, creativity and project management (Schalk & Bruning, 2014). What these competencies entail and the level expected from students at the end of the final year is described in the so called 'eindtermen' or final learning goals. These learning goals are centrally determined by the Dutch ministry of culture, science and education and therefore the same for all schools offering O&O (Schalk & Bruning, 2014). To support teachers in developing a curriculum that leads their students to reach these learning goals, a guidance document called 'handreiking schoolexamen O&O' was developed (Schalk & Bruning, 2014). The relationship between the goal of

preparing students for beta-technical jobs and studies and the described eindtermen is however not explicit, explanations given about the eindtermen do not link back to the skills needed in studies and jobs, but instead to the skill needed to run O&O-projects (Schalk & Bruning, 2014).

According to Visser (2023) the eindtermen en handreiking schoolexamen O&O do not provide sufficient clarity and guidance. O&O teachers indicate that they find the competencies complex and different teachers have different lists of competencies they focus on in their assessment (Visser, 2023). Additionally, different teachers interpret the competencies differently, causing there to be little consensus on what is expected of a student in order to graduate from O&O (Visser, 2023). Important to mention, however, is that at this moment the eindtermen are being revised by the so called 'vakvernieuwingscommissie' (NWV, n.d.)

### Assessment

Assessment in O&O education encompasses two distinct components: assessment of the product and assessment of the process (Lechner, 2012). Both components jointly contribute to the final assessment for the project, often resulting in a final grade for the project (Lechner, 2012). In the Netherlands ranges from a 1 to a 10 with a 5.5 being a passing grade.

The assessment of the product revolves around the quality of the final results of the project (Stichting Technasium, 2021). This part of the assessment is the same for every member of the team, as they are jointly responsible for their results (Stichting Technasium, 2021). In contrast, the assessment of the process involves an individual evaluation that considers the student's personal growth and their contribution to the project (Stichting Technasium, 2021).

The projects in the bovenbouw together make up the so called 'schoolexamen' that determine a student's final grade for O&O.

### Quality of O&O-assessment

In the following section, the previously described quality criteria of validity, reliability and fairness, are used to evaluate assessment in O&O.

From the literature, it does not become clear what assessment strategies teachers exactly use to come to a final assessment. Suggestions such as the use of portfolios and rubrics are given, but it is not clear how often these are used and how well they measure the intended constructs. Therefore, not much can be said about validity of assessments in O&O.

From the literature described it seems that the reliability of the assessment in O&O is low, because teachers do not agree on assessment criteria, their assessment strategies most likely differ significantly. Students have indeed expressed that they experience this (Visser, 2023).

Not much information about the fairness of O&O-assessments could be found in the literature. Therefore, no conclusions can be reached about this aspect.

### Study objectives

The main research question as described in the introduction was the following:

*What tool can be designed to effectively support O&O teachers in providing high-quality assessments in O&O?*

From the literature several requirements to good assessment have been identified. As shown in the model in Figure 1, there needs to be an alignment between Goals, Learning Activities and Assessment,

which is supported by a clear description of the constructs to be assessed. The clear description of these constructs is however difficult and potentially even undesirable for constructs related to personal and professional development, making a case for leaving in a subjective element. Regardless of the type of assessment, its quality can be described using the criteria of validity, reliability and fairness.

While a general description of O&O and assessment in O&O has been given, the literature search did not provide much clarity on the alignment of the different elements of the O&O-curriculum. Additionally, much remains unclear about the levels of validity, reliability and fairness in O&O-assessment, while available information gives rise to significant concerns.

To gain more clarity on the current state of assessment in O&O, the following research questions have been formulated:

1. How are O&O-students currently assessed? (Current practices)
2. What do students and teachers consider quality criteria for good assessment practices? (quality criteria)
3. What problems/obstacles do students and teacher currently face in the assessment of students? (Obstacles)
4. What solutions do already exist to address issues with the assessment in O&O? (Solutions)

To later support the development of a solution to the problems faced in O&O-assessment, following research questions are used:

5. What could an assessment of students in O&O look like? (Future Practices)
6. To what extent do the developed tools support teachers in assessment? (Supporting assessment)
7. To what extent do teachers consider the developed tools of added value? (Added value)

## Methods

To develop a tool to support teachers in giving high-quality assessment, this design project used methods from Human-Centred Design (HCD). In HCD, users are involved in the design process to make sure the results align with their needs and preferences (Steen, 2011). In this project, teachers actively participated to ensure that the solutions designed were valuable and applicable to them. O&O students were also involved to gain a more complete understanding of assessment in O&O. Furthermore, input from university teaching staff who teach a CBL course was sought to get an external perspective on assessment in O&O and to get inspiration from similar courses.

To better understand the current practices, quality criteria, obstacles and potential solutions in O&O-assessment (research questions 1-4), interviews were conducted with O&O teachers, O&O students and university teaching staff. Afterwards, a focus group with O&O teachers was organised to validate the interview results and to specifically focus on the development of potential solutions.

Based on the results from the interviews and focus group, the conceptual model in Figure 1 was updated and the research question was reframed. Using this as input, six concepts were designed (research question 5) and evaluated together with an O&O teacher (research questions 6-7). Input from this evaluation session was used to select three concepts and develop them further. These designs were then evaluated with another O&O-teacher (research questions 6-7), whose input was used to select and develop one final design.

### Ethical considerations

Qualitative research methodologies that require the researcher's interpretation of results, plays a major role in this research. Therefore, transparency regarding my background and potential biases as the researcher is essential. Given my experience as an O&O teacher, I possess an intimate familiarity with the topic of this research. While this familiarity is advantageous due to my in-depth knowledge of the field, it also introduces the potential for bias, as I may unintentionally favour my own perspective in the analysis and design process. To ensure that the research and the design remains applicable and relevant to a broader audience of O&O teachers, while using the advantages of my background knowledge, deliberate measures have been taken:

1. During the interviews and the focus group, I deliberately minimized my personal input, prioritizing capturing the perspectives of the participants;
2. I approached the analysis without preconceived expectations of the outcomes, ensuring an unbiased evaluation;
3. The input from interviews and the focus group served as the primary guide for the design process, while drawing inspiration from my personal understanding of O&O.
4. In the evaluation sessions, I emphasized co-creation, actively sharing insights with teachers. However, I ensured teachers had the opportunity to express their perspectives before offering my insights.
5. When finalizing the design, I relied on the specific feedback from evaluation sessions and the perspectives of participants gathered throughout the research, while also drawing upon my own experience.

To further ensure the ethical standards of this research, the ethical commission of the faculty Behaviour, Management and Social Sciences (BMS) from the University of Twente granted permission to conduct this research (request nr. 230072).

All participants were informed of the ethical consequences of this study through a comprehensive brochure. They subsequently provided informed consent by signing the respective consent form. In cases where participants were under 16 years of age, informed consent was also secured from their parents or legal guardians.

## Participants

In total 22 participants were involved in this research, nine teachers, ten students and three university teachers. To recruit O&O-teachers and O&O-students, general emails were sent to schools in the regional networks Drenthe, Groningen, Friesland, Gelderland, Midden-Nederland and Overijssel. Because O&O-schools collaborate and exchange materials in these regional networks, attempts were made to recruit participants from a variety of regional networks. Due to insufficient responses to these general emails, and in efforts to diversify the participant pool, personal contacts were used to recruit additional participants.

For the recruitment of university teaching staff, a general email was directed to all staff members within the Industrial Design Engineering (IDE) department at the University of Twente. This effort yielded two participants, with an additional participant recruited through personal connections. To participate as a university staff member, participants were required to have experience teaching a CBL course.

A total of nine interviews were conducted, involving 15 participants. One focus group was organized with six participants and two evaluation sessions involving two participants. All but one of the sessions (interview, focus group or evaluation) were conducted in Dutch, one of the interviews was conducted in English to suit the participants. The sessions were conducted in person, except for one interview and one evaluation session, which were conducted through videocall. An overview of the participants can be found in Table 1.

The participant codes shown in Table 1 are used in the results section to refer to participants. Quotes from the focus group were also captured, but it was not recorded from who the response came. Therefore, no codes are given to the participants of the focus group. Which classes the participants of the focus group teach was also not recorded.

Table 1: Overview of participants. Codes are used in the results section to refer to a participant.

Method	Participant	Code	Gender	School	Network	Class(es)	Language
<b>Interview 1</b>	Teacher 1	T1	M	1	Drenthe	OB	Dutch
<b>Interview 2</b>	Teacher 2	T2	M	1	Drenthe	BB	Dutch
<b>Interview 3</b>	Student 1	S1	F	1	Drenthe	4	Dutch
	Student 2	S2	F	1	Drenthe	4	Dutch
<b>Interview 4</b>	Student 3	S3	M	1	Drenthe	6	Dutch
<b>Interview 5</b>	Student 4	S4	F	1	Drenthe	3	Dutch
	Student 5	S5	M	1	Drenthe	3	Dutch
<b>Interview 6</b>	Student 6	S6	M	2	Drenthe	2	Dutch
<b>Interview 7</b>	Student 7	S7	M	2	Drenthe	2	Dutch
	Student 8	S8	M	2	Drenthe	2	Dutch
<b>Interview 8</b>	Student 9	S9	M	2	Drenthe	2	Dutch
	Student 10	S10	F	2	Drenthe	2	Dutch
<b>Interview 9</b>	University 1	U1	F	-	-	-	English

	University 2	U2	M	-	-	-	English
<b>Interview 10</b>	University 3	U3	M	-	-	-	Dutch
<b>Focus group</b>	Teacher 3	-	M	3	Overijssel	-	Dutch
	Teacher 4	-	M	3	Overijssel	-	Dutch
	Teacher 5	-	F	3	Overijssel	-	Dutch
	Teacher 6	-	M	4	Groningen	-	Dutch
	Teacher 7	-	F	4	Groningen	-	Dutch
<b>Evaluation 1</b>	Teacher 8	E1	F	5	Friesland	OB/BB	Dutch
<b>Evaluation 2</b>	Teacher 9	E2	M	2	Drenthe	OB	Dutch

## Procedure

In this section, descriptions of the procedures used for the interviews, focus group and evaluation sessions will be given.

## Interviews

The interviews were conducted from February to April of 2023 and lasted between 22 and 55 minutes, with an average of 40 minutes. All interviews were recorded using an audio recorder. The interviews with students and teachers were semi-structured and used a cardboard interview guide (Figure 2) with the following questions:

- What is assessed?
- How does assessment take place?
- Who is involved in assessment?
- Which tools are used in assessment?
- When does assessment take place?
- For what reason does assessment take place?
- How often does assessment take place?
- What are the characteristics of a good assessment?
- Vision: what is the reason for assessment?

The questions and design of the interview guide were inspired by the curricular spiderweb designed by Van den Akker (2003). The questions intended to approach assessment from different angles, providing a holistic understanding of assessment in O&O.



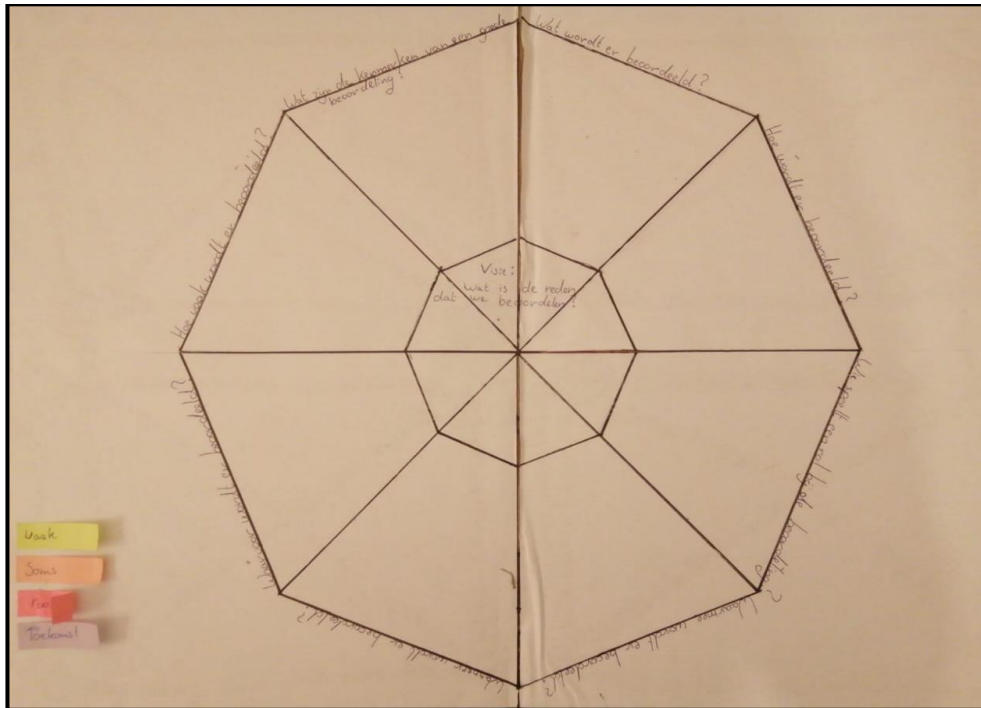


Figure 2: Cardboard interview guide with the questions: How often does assessment take place? How does assessment take place? Who is involved in assessment? Which tools are used in assessment? When does assessment take place? For what reason does assessment take place? How often does assessment take place? What are the characteristics of a good assessment? In the centre is written: Vision: what is the reason for assessment?

The interview guide was used to help participants visualise and organise their ideas. Participants were asked to talk out loud about the questions and to place post-its on the interview guide to summarize their responses. Additional and follow-up questions were asked when the researcher felt this was relevant. Four different colours of post-its were used to allow participants to indicate the following:

- Something that happens often (Yellow);
- Something that happens sometimes (Orange);
- Something that does not happen (Red);
- Something that the participant would like to happen in the future (Purple).

Figure 3 shows the interview guide after one of the interviews.

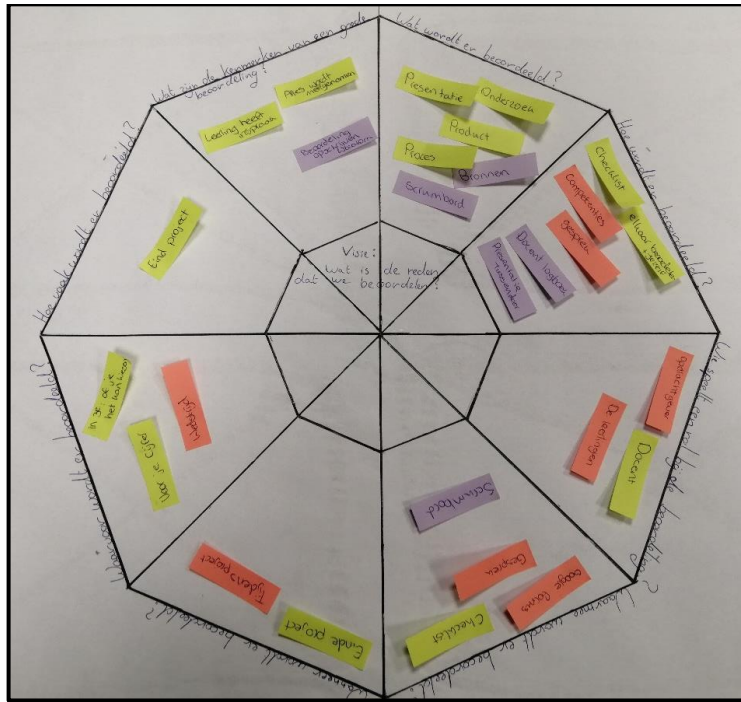


Figure 3: Interview guide after an interview with a participant.

The interviews with the university teaching staff were unstructured to allow for a more open discussion of their experiences with assessment in CBL. Their assessment strategies and advantages and disadvantages to these strategies were discussed. Participants were further encouraged to discuss quality criteria for assessment in CBL.

### Focus group

The focus group took place on the 16<sup>th</sup> of May 2023 at the Design Lab of the University of Twente and lasted for a total of three hours. To allow for participants to feel free to speak their mind, it was decided not to record the audio of the focus group. Instead, notes were taken by the researcher. Below an outline of the focus group is given, a summary of which can be found in Table 2. The section codes in this table are used in the results section to refer to a part of the focus group.

The focus group started with an icebreaker and introduction, to make participants feel comfortable and get them to start thinking about O&O. During the next activity, participants identified relevant stakeholders and their interests in assessment in O&O (Figure 4). Through discussing these diverse interests, participants then determined requirements for assessment in O&O.



Figure 4: Spider web diagram with stakeholders and their stakes used during the focus group

For each of these requirements, participants identified current and ideal situations (Figure 5), leading to a discussion about the most important differences.

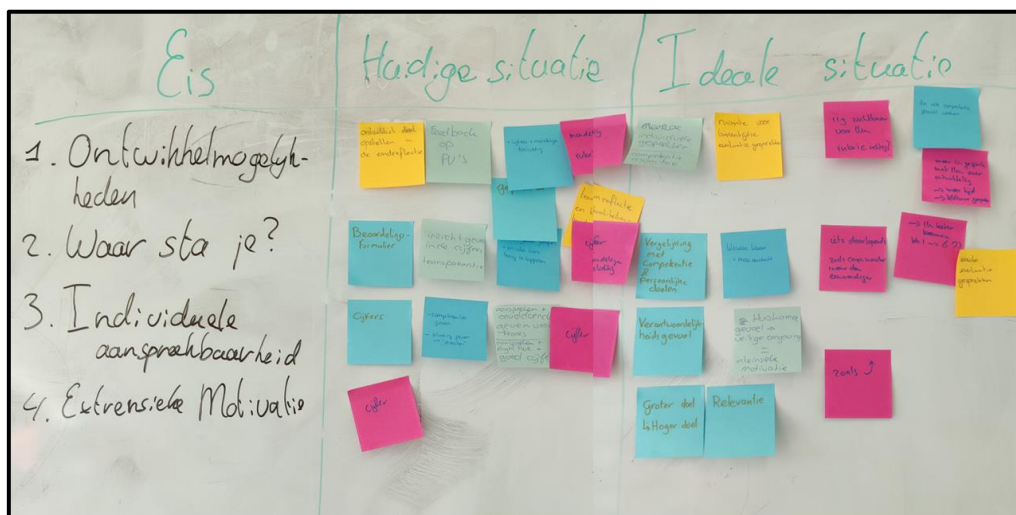


Figure 5: Table used to analyse requirement during focus group

Using these differences, participants were encouraged to come up with a problem statement through completing the sentence 'How could we ...?'. This technique was used to create many different problem statements, from which participants picked one to use for the remainder of the session.

To generate many potential solutions to the identified problem, a technique inspired by brainwriting (Paulus & Brown, 2003) was used. Participants drew or wrote down three ideas on a sheet of paper in five minutes. Afterwards participants handed their sheet to the next participant who drew or wrote down another three ideas. This continued on until all participants had written or drawn on all sheets, resulting in a total of 15 ideas per sheet (Figure 6). To familiarize the participants with all ideas, they were encouraged to look through all the sheets and pick their favourite solutions.

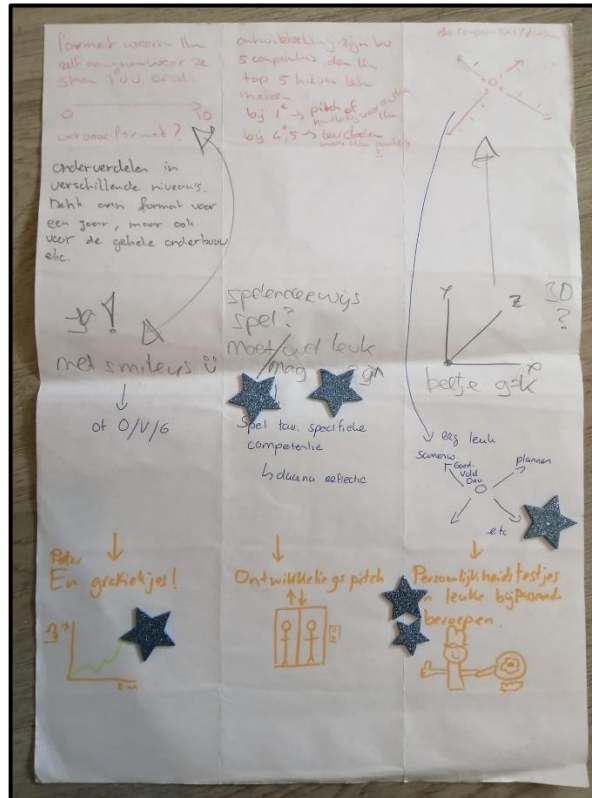


Figure 6: One of the sheet after the minddrawing activity

To add depth to the ideas, participants were asked to give a pitch about their favourite solution in groups of two or three. Finally, participants were asked to reflect on the focus group, giving their final thoughts on the session.

Table 2: Outline of the focus group. Codes are used in the results section to refer to a specific part of the focus group

Section	Code	Purpose	Time
Icebreaker and introduction	-	Making the participants feel comfortable Getting the participants to start thinking about O&O	25 min
Stakeholder analysis	FG-SA	Identifying stakeholders and their interests	25 min
Determining requirements	FG-DR	Identifying requirements of assessment in O&O	5 min
Requirements analysis	FG-RA	Identifying current approaches and future approaches to requirements Identifying most important gaps	20 min
Break	-	-	10 min
Determining problem statement	FG-PS	Generating potential problem statements Choosing one problem statement	15 min
Idea generation	FG-IG	Generating many potential solutions Choosing favourite solutions	40 min
Break	-	-	10 min
Idea generation	FG-IG	Pitching best ideas	10 min
Reflection and closing	-	Gathering final thoughts on focus group	20 min

## Evaluation sessions

Using the results from the interviews and focus group, an ideation process was used to come up with and select potential solutions, leading to a total of six initial concepts. The concepts were sketched on a flip-cart paper and evaluated during the first evaluation session, which took place on the 12<sup>th</sup> of Juli 2023. The session lasted a total of 1 hour and 28 minutes and was recorded using an audio recorder. After a short introduction of the research, the six concepts were discussed one by one using the following structure:

1. First impression of the concept based on the sketch without explanation;
2. Short explanation of the concept by the researcher;
3. Advantages of the concept;
4. Disadvantages of the concept;
5. Possible ways to improve the concept.

The intention of this evaluation session was to not just evaluate the concepts, but also develop them further. To encourage this, the teacher was asked to give a first impression of each concept before the explanation, already introducing a potentially different perspective in the conversation. Additionally, the teacher was invited to write and draw on the paper with the sketches. To stimulate this, the researcher also took notes on this paper (Figure 7).

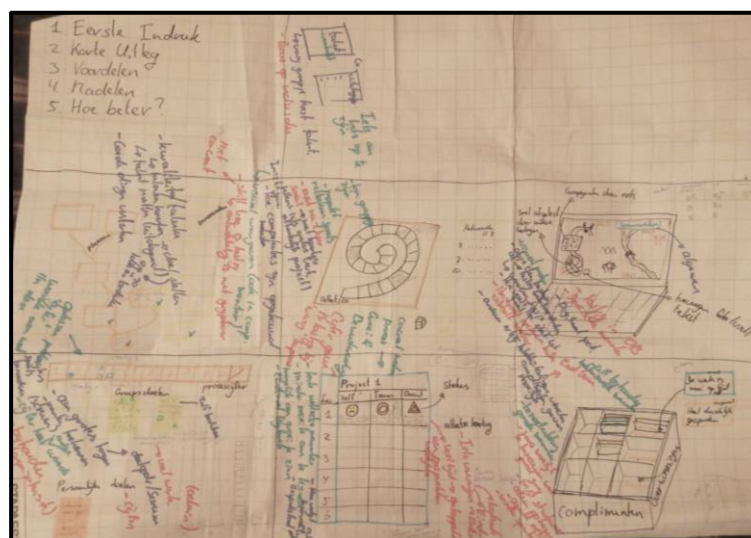


Figure 7: Paper with sketches used during the first evaluation session after the session. Both researcher and participant took notes and made sketches on this paper during the session

Using the input from the first evaluation session, three of the initial concepts were selected and developed further. These were discussed with another teacher during the second evaluation session. This session took place over videocall on the 13<sup>th</sup> of September 2023, lasted 40 minutes and was recorded.

At the beginning of the session, all three concepts were explained briefly, using drawings of the concepts. Afterwards, advantages and disadvantages of each concept were discussed in the form of an unstructured interview. It was decided to give one explanation of all three concepts and use the form of an unstructured interview to encourage the teacher to compare and contrast the different concepts instead of discussing them separately. The researcher guided the discussion to include all three concepts.

## Analysis

The recorded interviews were transcribed for further analysis using Amberscript (Amberscript, n.d.). These transcribed interviews, notes taken and artifacts produced during the focus group were coded using thematic analysis (Braun & Clarke, 2006). A deductive approach was used to code the responses to research questions 1 till 4, using the codes: current practices, quality criteria, obstacles and solutions.

For research question 1, the data extracts coded with 'current practices' were used to describe a coherent narrative on current assessment practices in O&O-education. For research questions 2 till 4, an inductive approach was used to identify and code specific quality criteria, obstacles and potential solutions. All of these codes were combined and interpreted, resulting in three overarching themes. The recorded evaluation sessions were coded deductively to find advantages, disadvantages and comments for each of the discussed concepts.

To support the coding process, the program ATLAS.ti was used (ATLAS.ti, n.d.).



## Results interviews and focus group

In this chapter, the results of the interviews with O&O-students, O&O-teachers and university staff and the results of the focus group with O&O-teachers are presented.

The analysis of results has led to the identification of three overarching themes: (1) quality assurance, (2), motivation and (3) student development, each representing a distinct assessment purpose in O&O. These themes are characterised by their own set of (1) current Practices (2) quality Criteria, (3) obstacles and (4) solutions. An overview of the information presented in this chapter can be found in appendices A till C.

After providing a detail examination of the three themes to answer research questions 1 till 4, a discussion on the tensions between these themes is presented. The analysis of the results indicated that the three assessment purposes often coexist in the assessment of O&O-students. This coexistence sometimes leads to tensions, as methods that benefit some of these purposes can cause issues with other assessment purposes. These tensions are outlined in final section of this chapter.

References to individual participants are made through their respective participant codes, as provided in Table 1. For the quotes extracted from source material or notes taken during the focus group, the codes are used to reference specific sections of the focus group, as indicated in Table 2. Any Dutch quotes and notes used in this chapter have been translated to English, with the original Dutch text included in footnotes. Throughout the chapter, models based on interpretations of the results are provided to visually support the text.

### Quality Assurance

Quality assurance is one of the three identified assessment purposes in O&O. This purpose is characterized by the process of checking a student's mastery of a construct against some *norm*. These norms could be set in place for reasons such as qualification, licencing or institution monitoring. While this assessment purpose was elaborately discussed during the interviews, it was only briefly mentioned at the beginning of the focus group when discussing the importance of assessment for different stakeholders.

*Somewhere that improvement needs to be documented, for your exam.<sup>1</sup> (S3)*

*To me, it doesn't matter that much whether you pass or not. This is however necessary for quality assurance.<sup>2</sup> (U3)*

*School management: Insight in quality of education<sup>3</sup> (FG-SA)*

As described in the Theoretical Framework, a student's mastery of a construct can be inferred from their performance. Therefore, when assessment serves the purpose of quality assurance, the assessment checks whether the performance of a student meets the predefined norms, as illustrated in Figure 8.

---

<sup>1</sup> 'Ergens moet die vooruitgang worden gedocumenteerd, voor je examen'

<sup>2</sup> 'Of je slaagt of niet, dat maakt wat mij betreft niet zo uit. Dat is wel nodig voor de kwaliteitsborging'

<sup>3</sup> Inzicht in kwaliteit van het onderwijs

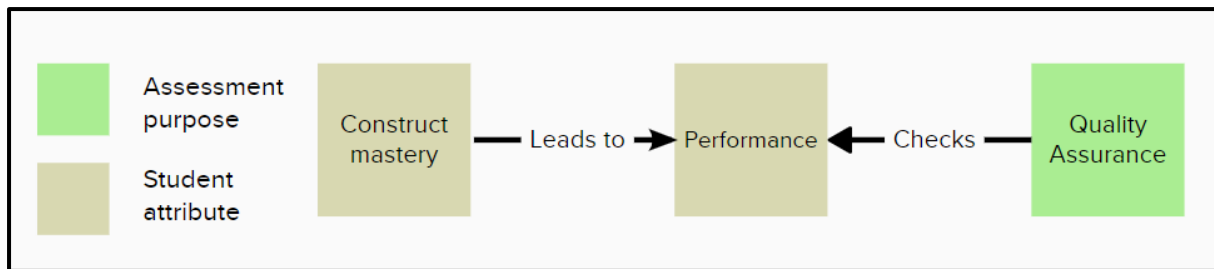


Figure 8: A schematic representation of the use of assessment for the purpose of quality assurance. The mastery of a construct is assumed to be represented in the performance of a student. The assessment checks if this performance meets predefined norms.

### Current practices

When participants were asked about current assessment practices in O&O, a notable distinction emerged between the participants from the two different schools. Nearly all participants from school 1 highlighted that assessment in O&O consists of two main components, the assessment of the product and the assessment of the process. However, none of the participants from school 2 brought up the assessment of the process when asked this question. Participants from school 2 did however mention elements that participants from school 1 classified as process assessment, such as attitude and behaviour.

For both the product and the process assessment, participants identified different elements that contribute to the assessment. For the product assessment these include the quality of the prototype, report and presentation, the inclusion of project-specific requirements and the team's ability to substantiate the choices made throughout the project. The process assessment, on the other hand, includes the student's progress toward achieving personal learning goals, behaviour during class, dedication to the project, the effectiveness of the collaboration and the adherence to the planning.

Both for the assessment of the product and the assessment of the process, the use of tools such as rubrics and checklists was often mentioned by the participants. One of the teachers mentioned that rubrics are most commonly used in the higher grades of O&O. According to him, in the lower grades of O&O, projects already have specific requirements, making checklists more appropriate.

As input for the process assessment, participants mentioned that students are often asked to write a reflection report and/or keep a log. Furthermore, teachers use their observations of students as input for the assessment.

*And if someone keeps chatting with another team over there, yeah, that really stands out. Well, I do take and those kinds of things into account in my assessment.<sup>4</sup> (T2)*

Students are additionally frequently asked to give their own perspective during assessment conversations, which can influence their and their peers' assessment.

<sup>4</sup> 'En als iemand continue dus bij [...] een ander team weer staat te ouwehoeren, ja, ja, dat valt echt wel op. Nou dan, en dat soort dingen neem ik dan eigenlijk wel mee [...] in mijn beoordeling.'



*With us, we are also asked for our opinion, like: 'suppose I take this part of your process, how would you then assess yourself on that'.<sup>5</sup> (S3)*

For the assessment of the product, the participants indicated that besides the teacher, other assessors are often involved in the assessment as well. These include the teacher, the client, the technical assistant (TOA), and sometimes an external expert. Students indicated that they appreciate the involvement of other professionals, as they believe it reduces teacher bias.

*A teacher and a client can also differ in terms of what they find beautiful. So, if you have, let's say, two sides; suppose the teacher doesn't find it so beautiful and would have given a lower grade, but because the client also reviews it and finds it beautiful, then it can still work out positively.<sup>6</sup> (S5)*

One of the teachers indicated that in the end, he is the one to interpret and decides the final assessment.

*That I just ask and then I get a certain impression. Then it's up to me how I'm processing it.<sup>7</sup> (T2)*

### Quality Criteria

During the interviews, participants discussed three different types of assessment norms, arguing that some are more suitable for assessment in O&O than others. Some participants emphasised accurately representing the student's level of competence, a standards-based norm. In the case of O&O this is usually reflected in the quality of the report, presentation or final product. The following quote is the answer of a student to the question whether someone who improved him/herself greatly should get a higher grade.

*No, letting the grade depend on the amount of improvement is not fair. This does not indicate the level, that is the way it is. You improved yourself, but this is not your level.<sup>8</sup> (S2)*

In this quote, the student emphasized the importance of the grade indicating the 'level' of the student. Conversely, some participants stressed that in O&O, a growth-based norm is more suitable. The following quote shows a student comparing O&O with other courses:

*And if you don't do anything, if you're as good at something and you keep doing it, for example in other subjects, then you will definitely get the same grades, and I*

---

<sup>5</sup> 'Bij ons wordt er ook naar onze mening gevraagd van: 'stel nou, ik pak dit onderdeel van jullie proces, hoe zou je dan jezelf daarop beoordelen?''

<sup>6</sup> 'Een docent en een opdrachtgever kunnen ook qua wat ze mooi vinden, ook verschillen. [...] Dus, als je dan ja, twee kanten hebt, stel de docent [...] vindt het niet zo mooi en die zou dan wat lager cijfer hebben gegeven, maar omdat de opdrachtgever er nog overheen komt en die vindt juist wel mooi, dan kan het net nog.'

<sup>7</sup> 'Kijk dat ik gewoon eens vraag en dan krijg ik dus wel een bepaalde indruk en [...] dan is het aan mij van hoe, als ik dat eigenlijk ga verwerken.'

<sup>8</sup> 'Nee, [cijfers af laten hangen van hoeveelheid vooruitgang] ook niet eerlijk, vind ik. [...] dit is niet het niveau, dat is dat is nu eenmaal zo. Je hebt vooruitgang gemaakt, maar het is niet je niveau.'

*also don't think should be allowed in O&O, because then you're not encouraged to be challenged.<sup>9</sup> (S3)*

Lastly, multiple students emphasized the importance of accurately reflecting a student's dedication to the project, an exertion-based norm. Students specifically argued that teammates that put in less effort should get a lower grade. Some students even argued that the assessment would be unfair if they receive a low grade despite putting in significant time and effort:

*If I think that I tried my best and my teachers gives me a four, I don't find that fair. With O&O that is, for physics that would be different.<sup>10</sup> (S3)*

Interestingly, for both the focus on development and the focus on effort, multiple participants from all three groups, indicated that this manner of assessment is distinct for O&O (or CBL) and that other courses (should) focus just on assessing the quality of the results.

Beside different assessment norms, participants also discussed the clarity of assessment criteria. Several participants, both teachers and students, pointed out that it is important that the assessment criteria are clear, to help students understand the rationale behind their assessment.

*If a student has a failing grade and doesn't understand why, then it's also not a good assessment.<sup>11</sup> (S4)*

One of the teachers mentioned that rubrics are particularly suitable for communicating these assessment criteria.

*An advantage of rubrics is that students can also see which requirements their product must meet.<sup>12</sup> (T2)*

While some participants emphasized the clarity of the assessment criteria, others pointed out the importance of taking a holistic perspective in assessment. This means that the assessment should not just be based on the detailed descriptions provided in the assessment rubrics, but look in depth at the performance of the student.

*The teacher should try to get to the core and not try to just put checkmarks, whether something is present or not.<sup>13</sup> (U3)*

While the use of a holistic perspective seems contradictory to the use of rubrics, one of the teachers combined both methods by providing a rubric without points. This rubric has detailed quality criteria, but no points are provided with the different components. This causes students to focus less on gathering points and allows the teacher to apply a holistic perspective in their assessment.

---

<sup>9</sup> 'En als je dus niks doet, als je zo [...] goed bent in iets en je blijft doen, dan bijvoorbeeld bij andere vakken, dan haal je sowieso dezelfde cijfers en ik vind ook niet dat dat mag bij O&O, want dan word je niet geprikkeld om uitgedaagd te worden.'

<sup>10</sup> 'Als [...] ik vind dat ik heel erg mijn best heb gedaan en mijn leraar geeft mij vier, dan vind ik niet dat het eerlijk is. Bij O&O dan, bij natuurkunde is een ander verhaal.'

<sup>11</sup> 'Als een leerling onvoldoende heeft en die snapt niet waarom, [...] dan is 't ook geen goeie beoordeling.'

<sup>12</sup> 'Een voordeel is van rubrics is dat leerlingen, [...] ook wel zien aan welke eisen hun product moet voldoen.'

<sup>13</sup> '[De docent moet proberen om] naar de kern te gaan en niet [proberen] om overall alle vinkjes maar te [...] zetten achter of iets aanwezig is of niet.'

*We used to assign points to the components, which actually creates a wrong incentive because then students write things down to get more points. What we have done, in any case, is not to put the points on the rubric anymore.<sup>14</sup> (T2)*

## Obstacles

During the interviews, participants discussed several problems they experienced with assessment in O&O. First of all, several participants expressed dissatisfaction with the current assessment criteria, critiquing them for being insufficiently clear, causing students to not know what aspects of a project are important to focus on.

*I would also like the requirements for the reports to be clearly available online, so that we know what we need to do.<sup>15</sup> (S9)*

*'That's the biggest challenge in challenge-based learning. [...] I don't think that at this moment there is clear guideline how we should do assessment in challenge-based learning' (U1)*

*Only nobody actually knows what the assessment is conducted with.<sup>16</sup> (S9)*

The lack of clarity in the assessment criteria may, in part, stem from the nature of O&O. This is because personal development is a fundamental component of O&O-education, which, according to some participants, is difficult to grasp.

*'And when you judge other things besides the disciplinary knowledge, like the progress that we said. You cannot really stick to numbers and you cannot also guarantee success.' (U2)*

Participants furthermore feel that the assessment is not always fair, with some students expressing strong sentiments regarding this issue. The following quote was made by a student who felt that she contributed more to the project than one of her teammates.

*We had, let's say, a 5,7, and she had something like almost an eight. And that didn't feel fair at all.<sup>17</sup> (S1)*

The participants disclosed several ways in which an assessment can be unfair, including students being dishonest about the contributions of their classmates, different clients giving different assessments and students profiting from the efforts of their classmates.

*In the previous project, two students did a lot, and one didn't do anything at all. And then in the end, you don't want to make someone look bad, so you start*

---

<sup>14</sup> 'Voorheen [...] gaven we punten op de onderdelen en dan krijg je een foute prikkel eigenlijk, want dan gaan leerlingen dingen opschrijven om meer punten te krijgen. [...] Wat we in elk geval hebben gedaan, is de punten niet meer op de rubriek zetten.'

<sup>15</sup> 'Ik zou ook de eisen van de verslagen dat die duidelijk online staan, dat we weten waar wat we moeten doen.'

<sup>16</sup> 'Alleen niemand weet waarmee er wordt beoordeeld eigenlijk.'

<sup>17</sup> 'Wij hadden zegmaar een 5,7. gekregen en zij had iets van bijna [...] een acht [...]. En dat voelde echt niet eerlijk.'

*writing: 'yes, they did this and that', but you feel guilty, it's not true. It's not fair, it wasn't fair.<sup>18</sup> (S1)*

Another way in which students feel that assessment can be unfair, is teachers not being aware of what is going on in the classroom. Because observations play an important role in assessment in O&O, it is important that teachers are noticing each of the students and are aware of how hard they worked. Both students and teachers pointed out this issue.

*Student 4: The teachers doesn't know how the process went.*

*Student 5: He does not have eyes everywhere, so to speak. He can't be involved in everything.<sup>19</sup>*

*It's sometimes a bit of a guessing game, you know. How much has someone actually done?<sup>20</sup> (T2)*

### Solutions

During the interviews, participants proposed several solutions for the identified obstacles. To improve the clarity of the assessment criteria to the students, the university staff suggested to make the rubrics together with the students. In this way the rubrics can be personalised, the students are more familiar with the criteria and they are likely to understand them better.

*'But another approach [...] was to make [the rubric] together with the students. So, you create learning objectives with the students, which means that the rubric will be based on those specific learning objectives.'* (U2)

To reduce problems with teachers not noticing everything, two solutions were put forward. For the first solution, students proposed that teachers keep a log as well. In this log, teachers could keep track of what stands out to them about the behaviour of certain students, making it easier to remember the behaviour of students throughout the classes, therefore making the final assessment simpler and more fair.

*That's why a teacher should also keep a logbook for each individual, and they don't have to write down everything there. But, you know, I notice that he is standing at a different group every time.<sup>21</sup> (S8)*

The second solution was students noting down their task divisions beforehand working on their projects and comparing this to the actual task division afterwards. In this way, it becomes more clear how much each team member contributed.

---

<sup>18</sup> 'Bij het vorige project, [...] deden twee [leerlingen] heel veel en eentje deed helemaal niks. En dan uiteindelijk aan het einde [...]. Je wil iemand niet zwak maken, dus dan ga je wel schrijven: ja, die heeft dit en dat gedaan, maar je krijgt een schuldgevoel, [...] het is niet waar. Het is niet eerlijk, het was niet eerlijk.'

<sup>19</sup> Leerling 4: 'De docent die weet niet hoe jullie proces is gegaan.'

Leerling 5: 'Die heeft niet overal ogen zegmaar. Die kan niet overal bij zijn.'

<sup>20</sup> '[Het] is toch [...] soms ook wel een beetje gissen van ja, hoeveel heeft nou iemand daadwerkelijk gedaan?'

<sup>21</sup> 'Daarom zou een docent ook een logboek moeten bijhouden per persoon, en hij hoeft niet alle daar alles op te schrijven. Maar goh, het valt me dat [...] hij iedere keer bij een ander groepje staat.'

*That a sort of list should be made, starting with a task distribution at the start, and then, after for example a week or two, what that person has achieved, how far they've come, and whether the task division is still the same.<sup>22</sup> (S2)*

## Summary

In this section Quality Assurance is brought forward as one of the three identified themes. From the interviews it became clear that O&O assessment currently consists out of a product and process part of the assessment, each with their own elements. Tools such as rubrics and checklist are used to support the assessment process and for the product part of the assessment several assessors are involved.

Participants mentioned three different norms for assessment in O&O: standards-based, growth-based and exertion-based norms. With the latter two specific to O&O/CBL according to some participants. Furthermore, clarity of assessment criteria and holistic assessments were mentioned as criteria for good assessment practices, with their apparent contradiction resolved by providing rubrics without points.

Discussed obstacles included dissatisfaction with the current assessment criteria and a feeling of unfair assessments. To improve the clarity of the assessment criteria, university staff suggested to make the rubrics together with the students. To improve the fairness of the assessment and specifically the problem of teachers not being aware of what is going on in the classroom, participants suggested to either let teachers keep a log or to let students keep track of their task divisions.

## Notes

*Monitoring of development<sup>23</sup> (Parents, Teachers, School management) (FG-SA)*

While the monitoring of development is relevant for quality assurance, it was not explicitly mentioned during the interviews. This monitoring is especially relevant when focusing on development as the content of the assessment, in order to assess how much a student has developed, it is necessary to the development of the student.

*The average of what others thought of you and what you thought of yourself, that then becomes your grade.<sup>24</sup> (S4)*

## Motivation

While assessment can be used to ensure the quality of a student's work, it can also play a crucial role in motivating students to excel on their projects. Assessment does, in this case, not cause students to master the constructs better, but instead stimulate them to try harder and in that way perform better, as depicted in Figure 9. Students either work harder, so that their competence is better shown in their performance, or their competence is increased through their hard work.

---

<sup>22</sup> 'Dat [...] gewoon een soort van lijst moet gemaakt worden van eerst een [...] taakverdeling aan het begin dan na bijvoorbeeld een week of twee [...] wat diegene [heeft] gedaan, hoe ver [die] is gekomen en of nog steeds dat de taakverdeling is.'

<sup>23</sup> Monitoren van ontwikkeling

<sup>24</sup> 'Het gemiddelde van wat anderen van je vonden en wat je zelf van jezelf vond, dat wordt dan je cijfer'

*You can like anything, but if it is not graded, you will set your priorities<sup>25</sup> (S3)*

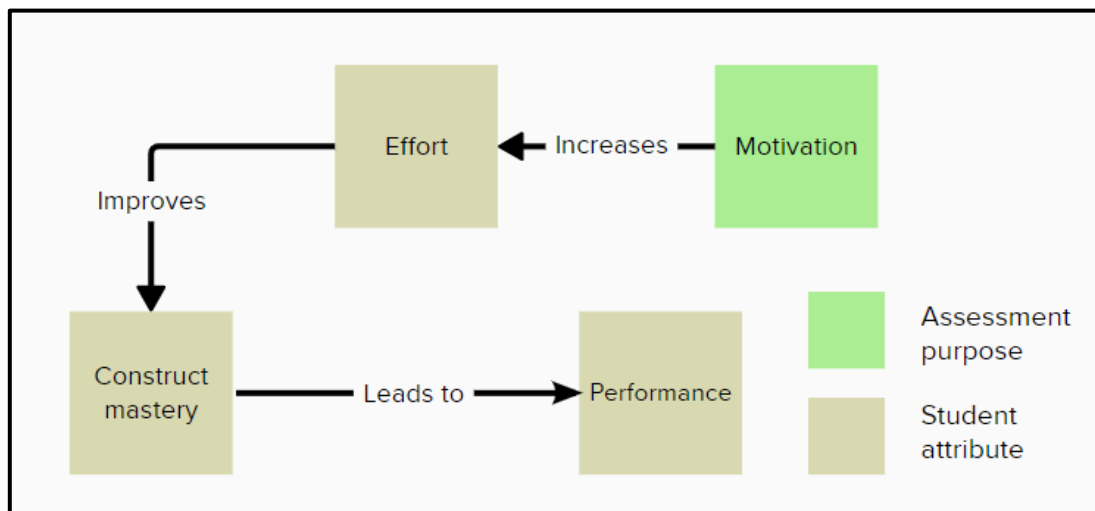


Figure 9: A schematic representation of the use of assessment for the purpose of motivation. The mastery of a construct is assumed to be represented in the performance of a student. The assessment stimulates the student to put in more effort, causing the performance to improve.

### Current practices

Grades currently play a major role in motivating students. Participants of the focus group even identified this as the main motivating force in O&O at the moment.

*I think that it is a kind of motivation. We naturally want a good grade, so that we will actually do it.<sup>26</sup> (S2)*

*'For me, grades were a driving force' (U2)*

### Quality Criteria

Two different perspectives were expressed by the participants for using assessment to motivate students. On the one hand, participants sought to increase the self-esteem of students and on the other hand to push students. When assessment is used to boost a student's self-esteem, it can serve as an affirmation of a student's efforts and accomplishments. For the two participants expressing this perspective, the aim was to instil a sense of pride in the students for their work and demonstrate how they have excelled.

*Yes, the reason we assess is a kind of confirmation. Ideally, I would give everyone a six and a half or a seven. I think, ah, you've done beautifully, very well, I'm satisfied with it.<sup>27</sup> (T1)*

On the other hand, some participants discussed the use of assessment as a tool to extrinsically motivate students to exert more effort and reach better results. These participants argued that students tend to focus their attention towards those things that are assessed.

<sup>25</sup> 'Je kan alles leuk vinden, maar als geen cijfer voor krijgt, dan ga je prioriteiten stellen.'

<sup>26</sup> 'Ik denk dat het een beetje een soort van motivatie is [...]. We willen natuurlijk een goed cijfer, dus dat we het wel echt gaan doen.'

<sup>27</sup> 'Ja, [de reden dat we beoordelen] is [een] soort bevestiging, [...] Het liefst zou ik iedereen zes en half of zeven geven. Ik denk van, ah je hebt het prachtig gedaan, hartstikke goed, ben d'r tevreden mee.'

*Because I think that if something is not assessed that students will look at it less seriously.<sup>28</sup> (S1)*

These two perspectives seem to be contradictory, as assessments designed towards building self-esteem tend to focus on positive outcomes, while assessment aimed at motivating students need to instil a sense of challenge to encourage students to exert effort.

As described students are currently mostly motivated through grades, however, participants expressed that in an ideal situation students would be intrinsically motivated. For this, students need to feel that O&O is personally relevant to them. To bring about this feeling, participants of the focus group suggested the following:

*Bigger/higher goals<sup>29</sup> (FG-RA)*

*A 'living room' feeling brought about by a safe environment<sup>30</sup> (FG-RA)*

*Giving students accountability<sup>31</sup> (FG-PS)*

*Motivating students to develop themselves<sup>32</sup> (FG-PS)*

Regarding the last point, one of the teachers pointed out the following:

*But if you're more focused on personal development, you dig a little deeper and the student also realizes: 'Hey, there's something to improve for me here, so I want to give it a try.' Which leads to more intrinsic motivation.<sup>33</sup> (T2)*

This shows that in order to foster intrinsic motivation, it is crucial for students engage in O&O in a way that is personally meaningful.

### Obstacles

The current emphasis on grades appears to cause students to adopt a shallow manner and show off those things a teacher wants to see, all in pursuit of higher grades. Concretely, this entails that students become dishonest and their independent thinking is hampered. Regarding the dishonesty, some students described instances where, in the pursuit of higher grades, they pretended to have exerted more effort than they actually did.

*We need to keep also track of how much time we spend working, we just exaggerate that a bit.<sup>34</sup> (S9)*

---

<sup>28</sup> 'Want ik denk dat [...] als je ergens niet op wordt beoordeeld dat het dan dat leerlingen er dan minder serieus naar gaan kijken.'

<sup>29</sup> Groter/Hoger doel

<sup>30</sup> Huiskamer gevoel → veilige omgeving

<sup>31</sup> Leerlingen eigenaarschap geven

<sup>32</sup> Leerlingen motiveren tot ontwikkelen

<sup>33</sup> Maar zit je meer op die persoonlijke ontwikkeling, dan graaf je iets dieper door en dan merkt die leerling ook zelf: 'hé, hier valt voor mij wat te verbeteren, dus dat wil ik ook gaan proberen.' Dus dan zit er meer de intrinsieke motivatie achter.

<sup>34</sup> 'We moeten ook bijhouden hoeveel tijd we bezig zijn, alleen dat dikken we een beetje aan.'

On the other hand, when students stop thinking independently, they focus on merely fulfilling predefined criteria, such as described in a rubric. This brings about a reduction in creativity and student initiative.

*If a grade is depending on it, then they are not directly involved with their personal development. They are more focussed on: 'What does my teacher want to hear from me, so that I get more points'.<sup>35</sup> (T2)*

### Solutions

A solution that was already mentioned to reduce the focus on grades, is to no longer add the points to the rubrics. To further combat the detrimental focus on grades, one of the teachers suggested the use of proper coaching techniques. Through these, students can become intrinsically motivated to work on developing themselves and take ownership of their learning.

*But if you focus more on that personal development, then you dig a little deeper and the student will notice for themselves: 'Hey, here I can improve myself and I want to try that', then there is more intrinsic motivation behind it.<sup>36</sup> (T2)*

### Summary

Ultimately motivating students is about either pushing them to put in more effort or building their self-esteem. This is currently pursued through the use of grades. However, this approach leads to issues with students being dishonest about their work and relying less on independently thinking. Ideally students would be driven by intrinsic motivating, which can be achieved by proper coaching techniques.

### Student Development

Due to the strong emphasis placed on student growth in O&O, an important assessment purpose is to facilitate the development of students. Student development was extensively discussed in both the interviews and the focus group. It was in fact the central point of interest during the focus group, as evident from the problem statement the participants decided to focus on:

*How can we get students to take a critical look at their own development?<sup>37</sup> (FG-PS)*

When assessment is used to support student development, it intends to give students the necessary input to improve their mastery of certain constructs. As an example, feedback on a student's presentation skills is intended to improve that student's mastery of presentation skills. This process is depicted in Figure 10.

---

<sup>35</sup> 'Er hangt een cijfer vanaf, dan zijn ze niet direct met hun persoonlijke ontwikkeling bezig, dan zijn ze meer bezig met van: 'Nou, wat wil de docent nou precies van me horen, zodat ik meer punten krijg'.'

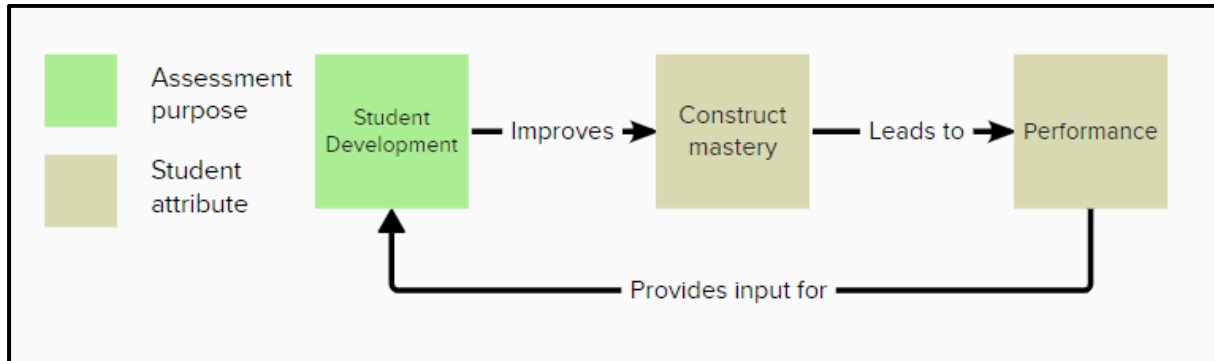
<sup>36</sup> 'Maar zit je meer op die persoonlijke ontwikkeling, dan graaf je iets dieper door en dan merkt die leerling ook zelf: 'hé, hier valt voor mij wat te verbeteren, dus dat wil ik ook gaan proberen', dus dan zit er meer de intrinsieke motivatie achter.'

<sup>37</sup> Hoe kunnen we leerlingen kritisch naar hun eigen ontwikkeling laten kijken?



*Assessment should primarily be meaningful for the development of students, that they derive substantive value from it.<sup>38</sup> (U3)*

*We really want students to become more aware of themselves and the choices they make.<sup>39</sup> (T2)*



*Figure 10: A schematic representation of the use of assessment for the purpose of student development. The assessment intends to improve the construct mastery of the student. The mastery of a construct is assumed to be represented in the performance of a student.*

While this focus on Student development can have a direct positive impact on the students and their projects through improved construct mastery, it also equips student with the necessary skills to face situations later in life. Both aims were discussed in the focus group and the interviews.

*So that they gain more insight into themselves, better know how to reflect on what they are doing. This is ultimately also beneficial for the team and the product.<sup>40</sup> (T2)*

*I think that this is preparation for, let's say, when you go to university, that O&O kind of encourages you to, when you encounter problems in life, how do you deal with them?<sup>41</sup> (S1)*

### Current practices

Through coaching conversations held throughout the lessons, assessment currently plays a continuous role in O&O. In these conversations, teachers give students input to facilitate reflection, aiding their growth.

*Every week/class, I take a moment to visit the teams and discuss their progress with them.<sup>42</sup> (T2)*

<sup>38</sup> 'Waarbij assessment vooral meaningful moet zijn [...] voor de ontwikkeling van [...] studenten, dat je er inhoudelijk wat aan hebt'

<sup>39</sup> 'Wij willen heel graag dat leerlingen bewuster worden van zichzelf en van de keuzes die ze maken'

<sup>40</sup> 'Zodat ze meer inzicht in zichzelf krijgen, beter weten te reflecteren op wat ze aan het doen zijn en dat uiteindelijk ook ten gunste is van het team en het product.'

<sup>41</sup> 'Ik denk dat dit [...] voorbereiding is voor zeg maar als je naar de universiteit gaat, dat O&O een beetje jou stimuleert van [...] als jij problemen tegenkomt in het leven, hoe ga je daarmee om?'

<sup>42</sup> Elke week/les even een rondje langs de teams en met ze in gesprek over de voortgang.

*Meanwhile, they look at, like, what's going well and what's going wrong? How can we further assist you, and whether it can be improved.<sup>43</sup> (S2)*

These conversations do not only take place throughout the lessons, but specifically at the end of the project as well, to review and discuss the assessment students received.

*He does review it, and then you sit in front of him and he reviews your report and, I don't know, your product.<sup>44</sup> (S10)*

Students can also make use of the so-called 'Competentiemonitor' to work on their personal development. This is a tool designed by Stichting Technasium (Stichting Technasium, 2021) to facilitate reflection on the Technasium competencies and assist in defining a path for self-improvement. Additionally, it can serve as a conversation tool for teachers.

*Such a person must be willing to realize that it's actually for their own benefit, and I believe that's how you should approach such a conversation with the Competentiemonitor.<sup>45</sup> (T1)*

### Quality Criteria

To encourage students to develop themselves, some participants emphasized the importance of personalised learning goals. This enables students to focus on those areas that are most beneficial to their growth. As previously mentioned, students furthermore tend to take greater ownership over such goals, causing them to be intrinsically motivated to engage with them.

*So as a teacher, you're constantly looking out for that, so to speak, waiting for when something comes up that the student really can work on, because the student also realizes it's beneficial.<sup>46</sup> (T2)*

However, one of the teachers pointed out that it is not possible for these learning goals to be completely personalized either, as students are ultimately required to show their proficiency in the learning goals outline in the 'eindtermen' (see Theoretical Framework). This teacher therefore expressed an intention to integrate these goals more and stimulate students to work on them.

*Yes, what I do want, is to incorporate more of the Technasium competences into that, to work with students on that a bit more.<sup>47</sup> (T2)*

To allow students to work towards the learning goals, opportunities for growth need to be created. As development is unpredictable according to the participants, creating these opportunities is not

---

<sup>43</sup> 'Ondertussen wordt er wel een beetje gekeken van ja, wat gaat er nou goed en fout? Waar kunnen we je verder mee helpen en ja, of het beter kan.'

<sup>44</sup> 'Hij behandelt het wel en dan [...] zit je voor hem[...] en behandelt hij je verslag en weet ik veel je product.'

<sup>45</sup> 'Zo'n persoon moet [...] willen inzien dat het dus eigenlijk voor hemzelf is en zo moet je ook zo'n gesprek ingaan, denk ik, met de competentiemonitor.'

<sup>46</sup> 'Dus als docent ben je constant daarnaar aan het loeren, om het zo maar te zeggen: van wanneer komt er nou echt iets naar boven waar die leerling mee aan de slag gaat, omdat die leerling dan ook merkt dat ie er wat aan heeft'

<sup>47</sup> 'Ja, wat ik wel wil is meer de Technasium competenties daar nog in kwijt, meer met leerlingen, daar wat meer een beetje mee leren omgaan.'

straightforward. However, the participants described circumstances that make growth more likely to happen.

*You can facilitate growth, you can create the conditions, you can ensure that there's reflection, that you have the time to turn those experiences into growth, but it's not possible to completely predict where it will be.<sup>48</sup> (U3)*

One of the conditions described is that students need sufficient high-quality feedback to know how they can develop themselves.

*That it is primarily meaningful and provides feedback that you can work with. But I often see that this is not always the case, that it is sometimes so ritualized, with assessment using rubrics and figures, where you can provide a judgment, but it doesn't really provide a stimulus for development in students.<sup>49</sup> (U3)*

Related to this, explanations of the assessment allow students to develop their own sense of quality and learn from their mistakes. One of the teachers therefore indicated that he always has a conversation with students to discuss their results. Participants felt that these assessment conversations significantly contribute to the development of the students.

*If I just fill in the rubric and give it back, then they have to figure out for themselves: yes, where does that come from in what I did and how I worked? Yes, I need to have a conversation with them about that. [...] And then you also need to point out the observations you've made or the documentation you've received, saying: 'Yes, I haven't seen this and this, or that's still missing, or you've developed very well in this and that.' So, I think you have to be able to indicate those things individually for each student; otherwise, they can't grow in those things.<sup>50</sup> (T2)*

Students do also indicate that this helps them to understand what they still can work on.

*But I actually understood it myself after he said that, when I actually looked at it afterward, I didn't really find it that nice.<sup>51</sup> (S6)*

---

<sup>48</sup> 'Je kan [groei] faciliteren je kan de condities creëren, je kan zorgen dat de reflectie erop zit, dat je de tijd hebt om van de ervaringen om daar groei van te maken, maar dat is niet van tevoren ja helemaal in te dekken waar die zal [...] zitten.'

<sup>49</sup> 'Dat het vooral meaningful is en en feedback oplevert waar je wat mee [...] kunt. Maar dat ik zie wel ook vaak dat dat niet altijd het geval is, dat het soms zo geritualiseerd is, die assessment met [...] rubrics en met figuurtjes dat je het wel een oordeel kan geven, maar dat het niet echt een [...] stimulans tot ontwikkeling geeft bij studenten.'

<sup>50</sup> 'Als ik alleen maar de rubric invul en ik geef dat terug, dan moeten ze zelf maar uitvogelen van: ja, waar ligt dat dan aan in wat ik heb gedaan en hoe ik heb gewerkt? Ja, dat daar moet ik met ze in gesprek. [...] En dan moet je zelf ook de observaties die je hebt gedaan of wat je binnen hebt gekregen aan documentatie, dan zul je moeten zeggen: 'Ja, ik heb dit en dit niet gezien, of dat of dat mist nog of daar en daar heb je heel goed in ontwikkeld.' Dus die zaken die moet je denk ik per leerling wel ook individueel kunnen aangeven, anders kan diegene niet daarin groeien.'

<sup>51</sup> 'Maar ik snapte het eigenlijk ook wel nadat ie dat gezegd had, [...], ik vond toen ik daar eigenlijk ook naar keek, uiteindelijk daarna, vond ik het eigenlijk ook niet zo heel mooi.'

Something that was discussed during the focus group, but not during the interviews, was to focus on the talents of students. This causes students to be aware of their own strengths and those of their peers, so they know what each student can contribute to the project.

Lastly, the participants expressed that the manner of assessing should depend on the age and maturity of the students. One of the students mentioned that older students have developed a more professional attitude, not always striving to get a grade at all costs. This allows teachers to have a more genuine discussion with them, as illustrated in the following quote. This quote was made in the context of a teacher asking a student what grade they think they deserve.

*I know that two years ago, I simply would have said: 'All or nothing, give me a nine.' Why? Because the worst that could happen is that he says: 'No, we won't do that,' we'll do something lower.'<sup>52</sup> (S3)*

### Obstacles

Throughout the interviews and focus group, participants expressed several obstacles for student development.

First and foremost, one of the teachers expressed concern about students placing insufficient emphasis on their development. This causes students to mostly focus on developing a product instead of using the opportunities provided to develop themselves.

*I of course also hear students say: 'I find the product more important, not the process'. They really want to create and do things. It is sometimes easier to come up with a solution than to look at yourself.'<sup>53</sup> (T2)*

Furthermore, as expressed before, students need sufficient high-quality feedback. According to some students this is currently not the case.

*Right now, we're not getting any feedback, so we don't know what we're doing wrong.'<sup>54</sup> (S9)*

Teachers pointed out that one of challenges in providing sufficient high-quality feedback is the substantial amount of time it takes.

*For those individual conversations in that one class, that will take you four weeks.'<sup>55</sup> (T1)*

The lack of available time was an obstacle that was also continuously emphasised during the focus group. While teachers acknowledge the value of individual conversations with students, they express a significant lack of available time to engage in these.

---

<sup>52</sup> 'Ik weet dat ik zelf twee jaar geleden gewoon had gezegd: de dood of de gladiolen, doe mij maar een negen. Waarom? Het ergste wat je kan overkomen, dat ie zegt: nee, doen we niet, dan doen we wel wat lager.'

<sup>53</sup> 'Ik hoor hier natuurlijk ook wel steeds leerlingen zeggen van: ja, maar ik vind het product het belangrijkste en niet het proces. Ze willen heel graag dingen maken en doen. [...] Dat is soms makkelijker om een oplossing te bedenken dan naar jezelf te kijken.'

<sup>54</sup> 'Want we krijgen nu ook geen feedback dus we weten niet wat we verkeerd doen.'

<sup>55</sup> 'Voor die individuele gesprekken in die ene klas, die ben je vier weken bezig.'

## Solutions

Throughout the interviews and the focus group, participants developed several solutions to promote student development. Given that student development took a central role in the focus group, the majority of solutions were centred around this topic. Therefore, the many solutions brought forward are divided into general approaches. One of the conditions specifically brought forward in developing these solutions, was that the teachers preferred the solution to not be digital, because they feel students already spend significant time on their laptops.

To address the considerable issue of the lack of time teachers have for assessment, participants proposed several solutions. One of these was asking older O&O students to help out with assessment and coaching. In this way, teachers have more time to focus on issues that specifically require their attention.

*I found it nice to be able to ask questions to someone who had already done it, like, for example, an older O&O student. And maybe that's something you can consider in O&O. That way, that you normalize coaching in vwo, so your teacher doesn't have to answer a lot of questions.<sup>56</sup> (S3)*

Another approach was the use of comparative judgement. This is a technique used by one of the university teachers where the work of students is compared amongst each other and grades are added afterwards. According to this participant, this makes it easier to grade and therefore the grading process takes less time.

*'In comparative judgment, you get always two essays next to each other and you just say, this one is better than that one. And at the end [...] you get a list from the best to the worst and then you grade based on that. [...] It's easier to choose better things [...] if they see the whole picture.'* (U1)

Other proposed solutions were smaller classes, more class time and teaching the same students for multiple years. The last point was brought up, because getting to know the students currently takes up significant time as well.

To give students insight into what they have learned throughout the different projects, participants proposed to visualize the development of students through either the use of portfolios or some sort of skill tree. With such a tool, progress can be documented over time and specific memories can be stored. This tool was described as:

*A growth document in which student collect proof for specific competencies that are related to the process.<sup>57</sup> (IG)*

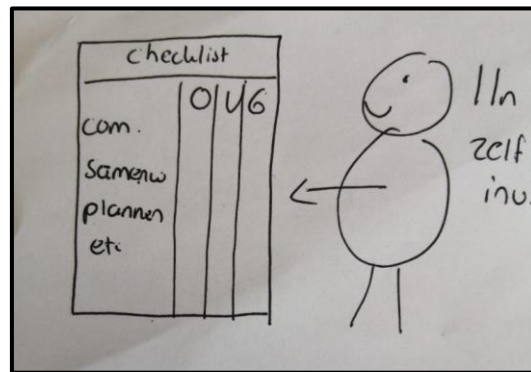
---

<sup>56</sup> 'Ik vond het fijn om vragen te kunnen stellen aan iemand die al keer het had gedaan, dus is dat bijvoorbeeld een oudere O&O leerling. [...] En misschien is dat wel iets in O&O waar je naar kan kijken. [...] Dat je dus in vwo coaching misschien iets normaliseert [...], waardoor je leraar niet en heel veel vragen moet beantwoorden.'

<sup>57</sup> Een groei document waarin leerlingen indicatoren (bewijzen) verzamelen voor specifieke competenties die betrekking hebben op het proces.

*You can, I think, define the competencies that need to be developed in O&O quite well, and from grade one to grade five and a half, you can indicate levels that the students should demonstrate they have achieved at some point.<sup>58</sup> (U3)*

Another tool that was proposed, was the use of self-assessment to help students reflect on where they are at, which can help them to determine future goals. An example of this can be seen in the drawing below (Figure 11).



*Figure 11: An example of the use of self-assessment to help students reflect on their competencies. The text reads: student fill in themselves (IG)*

The development of a learning line was brought up several times as well. A document like that outlines the learning goals for each year/project of O&O, making sure all learning goals are addressed. It was both proposed to make a general learning line or a personal one. In the latter case, the learning line is specific to the student.

*Making curriculum and every year focus on one of the aspects that contribute to a good process.<sup>59</sup> (IG)*

The use of explanations to develop an understanding of O&O was mentioned both during the interviews and the focus group. A specific example that was proposed was the use of superheroes to set an example for students. In the picture below the superhero 'Ears' is drawn, a superhero with very large ears that can listen very well (Figure 12). This superhero is meant to serve as inspiration for the students and to set an example.

<sup>58</sup> 'Je kunt de competenties die [bij O&O] ontwikkeld moeten worden, kun je denk ik best goed definiëren en je zou van klas één tot [...] vijf en half [...], kun je best levels aanduiden die die de leerlingen een keer moeten laten zien dat ze dat gehaald hebben.'

<sup>59</sup> Curriculum maken en elk jaar de focus leggen op 1 van de aspecten die bijdragen aan een goed proces.



Figure 12: An example of a tool for explaining competencies. This drawing of Ears depicts a superhero that can listen very well.

### Summary

Through coaching conversations and reflection exercises, assessment with the purpose of student development plays a prominent and continuous role in O&O. Students actively participate in their own development by formulating personalised learning goals during coaching conversations with teachers. However, one of the teachers expressed concerns about a lack of interest from students in their development.

Additionally, in shaping these learning goals, the O&O eindtermen should be considered, as they indicate what is expected of all O&O students. Participants suggested the development of a learning line to outline the learning goals to be addressed per year.

To facilitate the development of students, creating the right conditions for growth is essential. This includes students receiving sufficient high-quality feedback as input for their projects. It furthermore entails discussing explanations of the assessment with student and focussing on their talents. Finally, participants expressed that the assessment process should be tailored to the age and maturity of the student.

However, some students have reported receiving insufficient feedback, which might be attributed to teachers' time constraints, as providing high-quality feedback is a time-intensive task. To address this issue, participants suggested potential solutions, including involving older O&O students or using comparative judgement techniques.

Several other solution paths were developed as well. Participants proposed visualising student's progress throughout different projects to provide them with insights into their learning journey. They also suggested implementing self-assessment tools to encourage students' reflection on their own development. Lastly, participants advocated for the use of explanations to give students a better understanding of certain O&O practices.

### Tensions

All three of the purposes of assessment that are discussed above currently play a role in O&O education. Given that these purposes have different requirements, their simultaneous occurrence can give rise to tensions. The three tensions that were discussed by the participants are outlined below.

### Accurate grading

When assessing for the purpose of quality assurance, it is important that the performance of a student is accurately represented in the assessment. Students that perform better should get a more positive assessment. This accurate representation is less important when assessment is used for motivational purposes. This is well reflected in the following quote that was also discussed before:

*Yes, the reason we assess is a kind of confirmation. Ideally, I would give everyone a six and a half or a seven. I think, ah, you've done beautifully, very well, I'm satisfied with it.*<sup>60</sup> (T1)

A similar line of thought becomes apparent in a quote from one of the university teachers, related to the purpose of student development. It is an answer to the question whether a teacher should hold up a mirror to a student during an assessment conversation.

*Yes, however a mirror tries to reflect you as accurately as possible. Sometimes, we also need to use a funhouse mirror, which can either exaggerate the ugly things or highlight the beautiful things about someone, depending on how they see themselves.*<sup>61</sup> (U3)

In both of these cases the focus is less on giving an accurate representation of the performance of a student and more on using assessment as a tool.

### Teacher's role

In O&O, teachers play an important role during the projects, as they help students develop themselves through their coaching. However, by providing feedback and guidance, teachers can influence the performance of their students, causing the final assessment to not be an accurate representation of a students' capabilities. This is reflected in the following quote:

*You're dependant on your teacher. If you don't have an intensive or motivated O&O teacher, you can't get high grades either.*<sup>62</sup> (S3)

An interesting way to address this concern was proposed by one of the students. This student suggested that students should assess their teachers on their ability to help out students and find suitable projects as well:

*I actually think that students should be able to give an assessment to the teacher, because then a teacher can pay attention: 'Is this really a good assignment for the students?'*<sup>63</sup> (S1)

---

<sup>60</sup> 'Ja, [de reden dat we beoordelen] is [een] soort bevestiging, [...] Het liefst zou ik iedereen zes en een half of zeven geven. Ik denk van, ah je hebt het prachtig gedaan, hartstikke goed, ben d'r tevreden mee.'

<sup>61</sup> 'Ja, alleen een spiegel probeert je zo [...] precies mogelijk weer te geven. We moeten soms ook een beetje een lachspiegel gebruiken, dat je die wat uitvergroot of dat lelijke dingen of juist de mooie dingen van iemand afhankelijk van hoe ze zichzelf zien.'

<sup>62</sup> 'Je bent afhankelijk van dus je leraar. [...] Als je geen intensieve of gemotiveerde O&O leraar hebt, [...], kan je ook geen hoge cijfers halen.'

<sup>63</sup> 'Ik vind eigenlijk dat dat ook wel vanuit de leerlingen een beoordeling, naar de docent mogen geven. [...] want dan kan de docent ook er opletten: Is dit wel echt een goede opdracht voor de leerlingen?'



This does not immediately solve the problem of teachers influencing the performance of the students, it does however give teachers feedback on their actions.

### Working for grades

While grades can be a powerful tool motivate students, the focus on grades can cause them to focus less on their own development, therefore causing problems with the purpose of student development.

*Because there's a grade involved, they're not directly focused on their personal development. Instead, they're more concerned with what the teacher wants to hear so that they can get more points.<sup>64</sup> (T2)*

So instead, one of the university staff members decided to not assess the development of students in order to allow students to develop themselves in surprising ways and avoid regulation of growth.

*Then it's also about experiencing developments that you didn't expect.<sup>65</sup> (U3)*

*We hope there's some growth in it, but we're just afraid that as we further regulate, the growth may diminish, and it might become more of an exercise.<sup>66</sup> (U3)*

This idea of not assessing personal development seems to be in line with the idea of creating the right conditions for growth. According to this university teacher, allowing students to grow is easier when there is no external pressure to do so.

### Summary

In this section, three tensions between different assessment purposes were discussed:

- Accurate assessment: For quality assurance purposes the assessment needs to represent the performance of the student. For other purposes other factors play a role in determining the assessment;
- Teacher's role: Through their coaching, teachers play an important role in the O&O-projects of students, which causes them to influence the performance of the students;
- Working for grades: Students being motivated by grades can hinder their development.

---

<sup>64</sup> 'Want er hangt een cijfer vanaf, dan zijn ze niet-direct met hun persoonlijke ontwikkeling bezig, dan zijn ze meer bezig met van nou, wat wil de docent nou precies van me horen, zodat ik meer punten krijg.'

<sup>65</sup> 'Dan gaat het juist ook om dat je ontwikkelingen doormaakt die je die je niet verwacht had.'

<sup>66</sup> 'We hopen dat je dat er wat groei in zit, zijn alleen bang dat op het moment dat we verder gaan reguleren, dat er groei minder wordt en dat het een soort [...] exercitie gaat worden.'

## Reframing

In the Introduction and the Theoretical Framework, several challenges with assessment in O&O education are outlined. These sections point towards a general lack of consensus and clarity in the implementation of assessment strategies. In order to delve deeper into the current practices, quality criteria, obstacles and solutions related to assessment in O&O, interviews and a focus group were conducted and their results were analysed. Outcomes of these analyses do not only confirm some of the challenges previously outlined, but also provide a more in-depth understanding.

In this chapter, analysis of the results in combination with previously discussed literature serves as a basis for reframing the research question. This reframed question will subsequently guide the design process in the following chapter.

### Assessment purposes

From the analysis of the results three distinct assessment purpose were identified: (1) *quality assurance*, (2) *motivation* and (3) *student development*, each playing a unique role in O&O assessment. The three different assessment purposes are visually represented in a combined model (Figure 13).

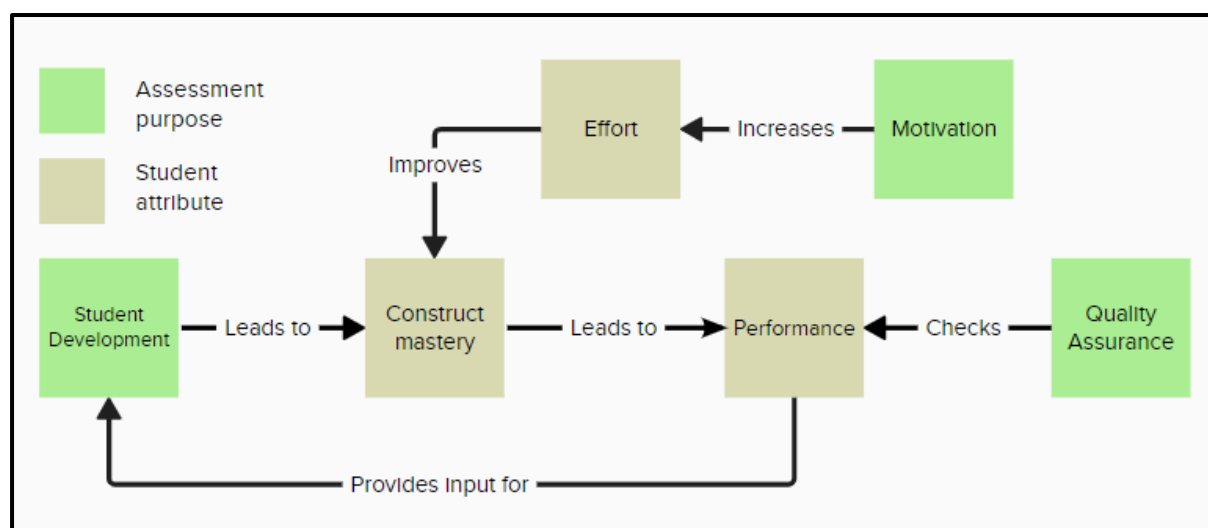


Figure 13: Combined model of the three purposes of assessment in O&O. Quality assurance checks the performance, while student development improvement the construct mastery of a student and motivation increases the effort a student puts into the project.

### Quality Assurance

When discussing the assessment of O&O students for the purpose of quality assurance, participants brought up three different norms: Standards-based, Growth-based and Effort-based. These norms are simultaneously and inconsistently applied in O&O assessment, particularly in the context of process assessment, which includes: effectiveness of the collaboration (Standards-based), progress toward achieving personal learning goals (Growth-based), as well as dedication to the project (Effort-based). At the same time, students have expressed confusion about the assessment criteria, a sentiment echoed in the research conducted by Visser (2023). This supports the suspicion, expressed in the Theoretical Framework, there exists a general lack of consensus and consistency in O&O assessment practices.

This lack of consensus and inconsistency poses significant issues, rendering the assessment both invalid and unreliable. The assessment's invalidity stems from absence of construct clarity, as no

statements about proper construct representation can be made when the construct to be assessed is unclear (AERA et al., 2014). Additionally the assessment is unreliable, since assessment standards vary across schools and teachers (AERA et al., 2014). This lack of validity and reliability is reflected in the responses of students during the interviews, when they criticise the assessment for being unfair. In this context 'unfair' is not used in a technical sense to imply bias or favourable treatment, rather it alludes to concerns related to the assessment's reliability and validity (AERA et al., 2014).

However, this does not negate the potential for the deliberate inclusion of subjective assessment practices, as highlighted by Schuwirth and Van der Vleuten (2012). The key distinction is that even subjective assessments require consensus and clarity in the assessment criteria. Once these criteria are well-defined, the interpretation of construct mastery is left to the subjective judgment of the assessor(s).

As previously mentioned, this subjective assessment could manifest through the jury model, as proposed by Hodge (2013). In fact, this model is already employed in the product assessment of O&O. Here, multiple stakeholders collaborate in assessing students' products. It is however difficult to implement a similar approach for the process assessment, as that would necessitate the continuous availability of additional assessors.

'The Standards' outlined both Standards-based and Growth-based norms in addition to Normative norms (AERA et al., 2014), which were not discussed by the participants. Conversely, the Effort-based norm identified by the participants, was not discussed in 'The Standards'. This is not surprising, given that the Effort-based norm fundamentally differs from the other norms, as it assesses a student's input or effort rather than their output or performance (see Figure 14).

However, both approaches align with the given definition of assessment as 'the process of first checking and then making decisions based on the extent to which students have mastered certain constructs.' As previously discussed, construct mastery cannot be empirically observed, but only approximated through observable behaviour. Under the assumption that increased effort correlates with better construct mastery, assessing effort is a viable means of approximating construct mastery, similar to using performance. Given the difficulties associated with performance assessment in O&O and other CBL courses, it is unsurprising that construct mastery is approached through an emphasis on effort. However, it is interesting to note that this approach is considered uncommon and, according to the participants, undesirable for other courses.

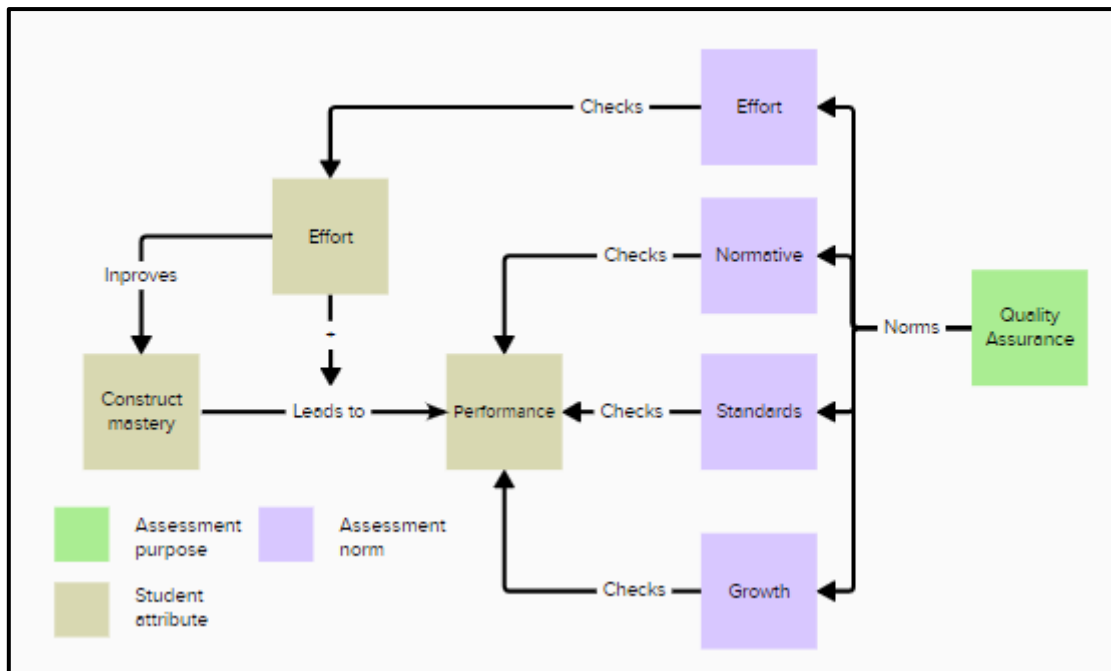


Figure 14: For quality assurance, four different norms have been identified from literature and empirical research. These norms assess the performance and effort of a student.

### Motivation

During the interviews and focus group, motivating students has been described as a justification for assessment. However, when revisiting the definition of assessment as given in the Theoretical Framework, it should be acknowledged that Motivation as an assessment purpose, does not align with the definition:

Assessment: The process of first checking and then making decisions based on the extent to which students have mastered certain constructs.

When assessment is used for motivational purposes, there is no process of checking and making decisions. Instead, the presence of the assessment is intended to stimulate students to invest effort in their projects. In essence, when assessment is conducted solely to motivate students, it can no longer be classified as a true assessment, as there is no interest in the outcomes of the assessment anymore. Consequently, it can be concluded that motivation should not be regarded as a primary purpose of assessment. This perspective is further supported by the absence of motivation as an assessment purpose in Newton's (2007) list of assessment purposes.

Nevertheless, the influence of assessment on motivation remains a relevant factor to consider. As motivation significantly affects the learning behaviours of students, it constitutes a form of washback (Green, 2007). Both the positive and negative consequences washback described by Baird et al. (2017) are reflected in the interviews as well. Assessments not only have the capacity to motivate students to invest effort into their projects, but can also cause issues such as dishonesty and reduced independent thinking.

### Student Development

Throughout the interviews and the focus group, it became clear that the development of students plays a major role in O&O. This is apparent in the assessment practices used, such as coaching conversations that intended to support student development. To give high-quality feedback in these

assessments, Hattie and Timperley (2007) argued that the feedback needs to be aligned with well-defined goals, clearly pointing out the gap between the current and desired state.

However, as expressed by some participants, a student's development in O&O is not about arriving at a pre-set destination. Rather, it entails students striving towards realizing the ideal version of themselves through a process of reflection and self-discovery. This deeply personal journey cannot be captured in clear and specific learning goals, making it impossible for teachers to align their feedback with these goals.

Nevertheless, teachers can, and according to the participants do, play an important role in guiding students' self-development. They should provide, as a university teacher called it 'meaningful feedback', which according to one of the participants in Visser's (2023) study is best guided by experience and intuition. Or as described by Hodge (2013) a 'holistic impression' provided by an experienced teacher. The central principle here is that assessment for student development should create 'space for growth'. Providing students those things they need to develop themselves.

Regarding learning goals, according to Black and Wiliam (2009), it is crucial that the teacher maintains some form of learning intentions, even if implicit, otherwise 'everything goes'. In the context of O&O, it appears most sensible for the learning intentions to be personalised for the student and emerge from discussions between teachers and students, while keeping in mind the pre-set learning goals outlined in the 'eindtermen'.

While this is a sound general principle, it may not be sufficient to aid teachers struggling with problems such as time investment and students not prioritising their development. Addressing these challenges requires specific guidelines and support.

### Updated Model

Using the reflections on the three assessments purposes as provided before, the model discussed in the Theoretical Framework (Figure 1) has been updated to include the additional relevant concepts (Figure 15).

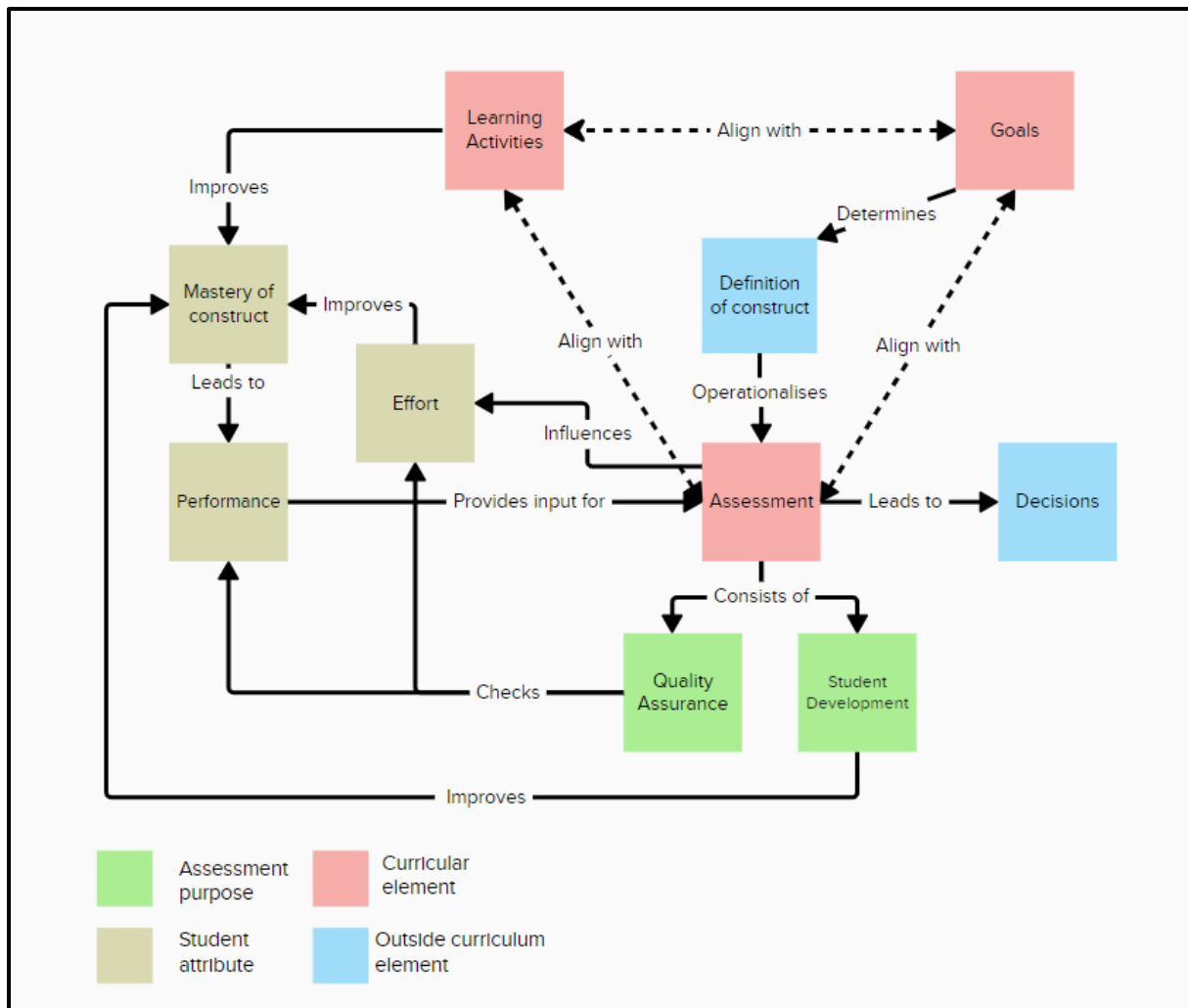


Figure 15: Updated model of assessment in education combining the model in Figure 1 with the model in Figure 13 to include both the results from the literature and the empirical findings.

The model in Figure 15 combines the model in Figure 1 with the model in Figure 13, with some notable differences. The concept 'construct' is now divided into a student attribute (construct mastery) and an outside curriculum element (definition of construct). This decision was made, because the definitions of the constructs are used to operationalise the assessment practices, while the mastery of the construct is an internal attribute of the student, leading to their performance.

Two assessment purposes are illustrated, quality assurance and student development. Motivation is no longer depicted as an assessment purpose, because of reasons outlined before. The motivational effect of assessment is however acknowledged by the arrow from assessment to effort.

### Reframing

Based on the considerations mentioned above, several directions for further exploration of the topic of assessment in O&O can be identified. First of all, for the purpose of quality assurance, there seems to be a need for improved clarity of learning goals and assessment strategies. This could strengthen the curricular alignment (see Figure 15), reduce validity and reliability issues and clarify assessment practices to students.

As the 'vakvernieuwingscommissie' (NWV, n.d.) is currently working on renewing the learning goals for the 'schoolexamen', it is expected that the refinement of these learning goals will be part of their ongoing efforts.

Another direction is to provide guidance towards assessment for student development. This seems a relevant direction, as assessment for student development was discussed at length during both the interviews and the focus group. Consequently, the research question has been reformulated as follows:

*What assessment tool can be designed to support teachers in facilitating and enhancing student development?*

Drawing upon insights from the literature, interviews, and focus group discussions, this research question will serve to guide the design process outlined in the next chapter.

## Design

In this chapter the design process will be outlined, describing the initial concepts, the first evaluation session, the concepts chosen following this session, the second evaluation session and ultimately the final design.

### Initial concepts

Initial ideas were generated using the brainwriting technique also used in the focus group (see Methods) (Paulus & Brown, 2003). To initiate this process, the prompt, "How can development be stimulated?" was used as a starting point. Additional ideas were elicited through a metaphor exercise. These ideas were subsequently compared and contracted amongst themselves and the ideas generated by participants during the interviews and focus group, resulting in the identification of general solution directions.

From this pool of ideas, the solution directions that had the greatest potential to align with the research question were selected, ultimately leading to the development of the six concepts outlined in this section. For each of these concepts, the main obstacle they aim to address is discussed alongside the underlying principles that guide their design.

These initial concepts are presented as rough sketches and solution directions instead of fully worked out concepts. This choice was made deliberately to allow for brainstorming opportunities during the first evaluation session.

### Challenge Track

The Challenge Track is based around the idea of creating space for growth discussed in the interviews. The focus is therefore not on assessing the unpredictable nature of growth, but rather on what actions a student has taken to allow for growth to happen.

Using the Challenge Track, students complete challenges during the project to earn points. The challenges are formulated in such a way that they encourage students to develop themselves. challenges can be either for the whole group or for individual students and teachers can decide if they let students pick the challenges themselves or they give them out randomly. Each of these challenges is worth a point and the number of points directly translates to the grade students get. A point tracker can be used to keep track of the points students have earned (Figure 16).

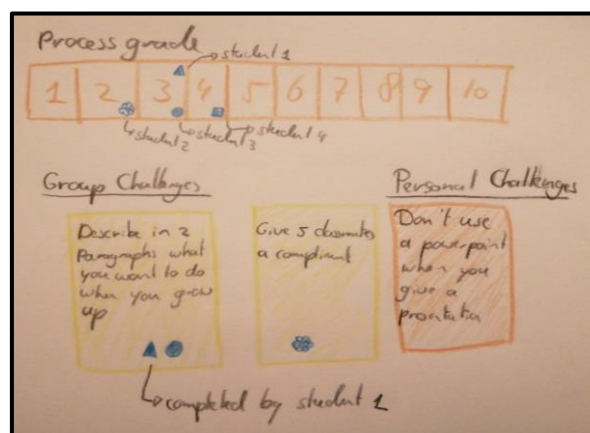


Figure 16: Sketch of design for the Challenge Track



This concept can be used to clarify the assessment criteria. Because of the set-up with challenges and points, it becomes immediately apparent to students what they need to do to improve their grade.

### Mood Card

The Mood Card is intended to simplify the reflection process and allow peers and teachers to easily give feedback. In this way, the concern of 'insufficient feedback' can be addressed. On the Mood Card, students, peers and teachers can place stickers during the lessons, indicating how they are feeling. Under the stickers, the student is encouraged to write a short reflection on why that sticker was placed there. Multiple stickers can be placed during each class, allowing students to show progress: e.g. at first they might feel demotivated, but after a while they are feeling better (Figure 17). In this way, the Mood Card can be used to visualise the developmental process of the students.

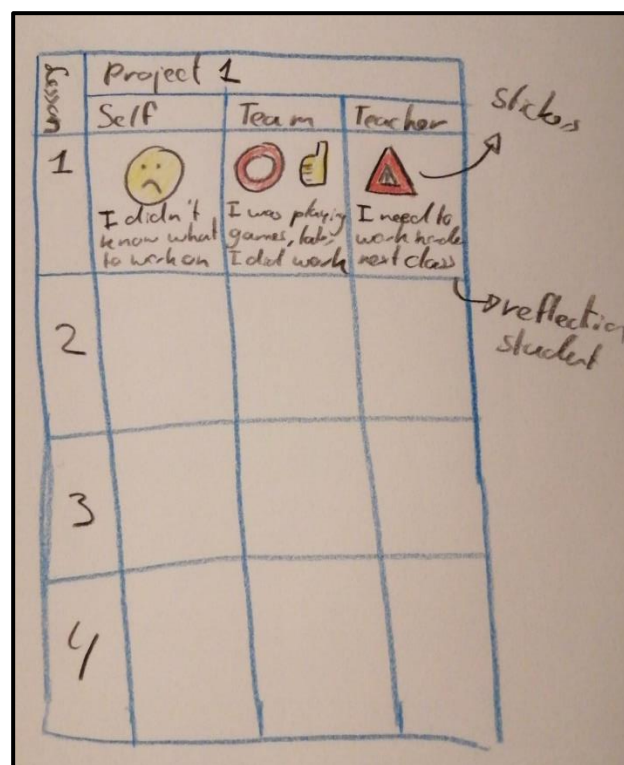


Figure 17: Sketch of design for the Mood Card

### Trophy box

The Trophy Box is intended to help students see their projects in a positive light and focus on their victories and proud moments. Though this it becomes clear where each student's talents lie. Each of the students has their own box with nine slots representing different competencies (e.g., Team work, Planning, Presenting). Peers, teachers and students themselves can add Compliment Cards to this box and fill it over the years. Because the cards for the different groups have different colours, it can easily be seen who gives compliments about what competency (Figure 18). By ritualising and simplifying the feedback process, the issues with a lack of feedback can be addressed.

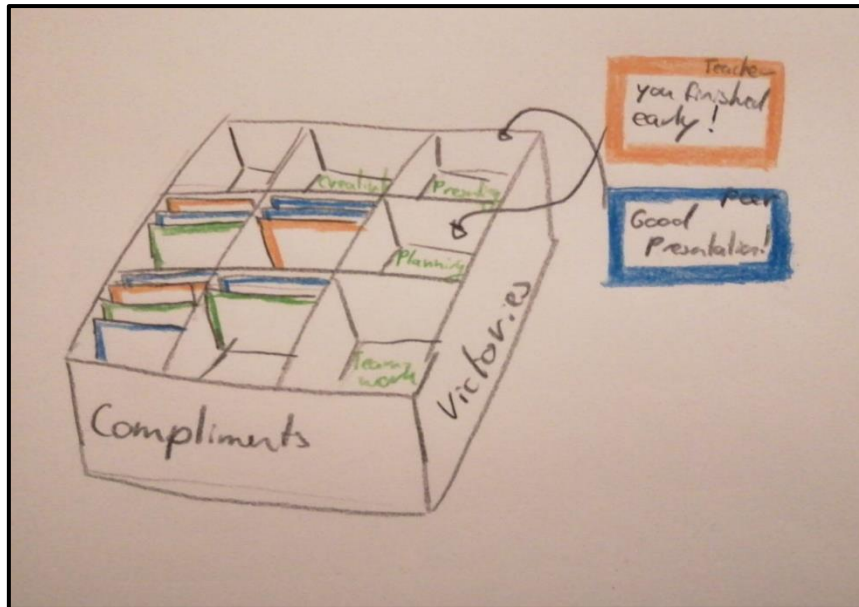


Figure 18: Sketch of design for the Trophy Box

### Journey Map

Obstacle: Intangible personal assessment

Underlying principles: visualising development

The Journey Map uses the principle of visualising the development of students through the use of a digital portfolio. The Journey Map is a visual representation of the obstacles and victories a student encounters in their competency development over the years.

For each of the Competencies, students visualise their journey using pre-made objects. Students can for example use a tree trunk on the road when they feel like they are stuck. To elaborate on this, students can add pictures, videos or text and they can ask teachers and peers to add comments or feedback as well (Figure 19).

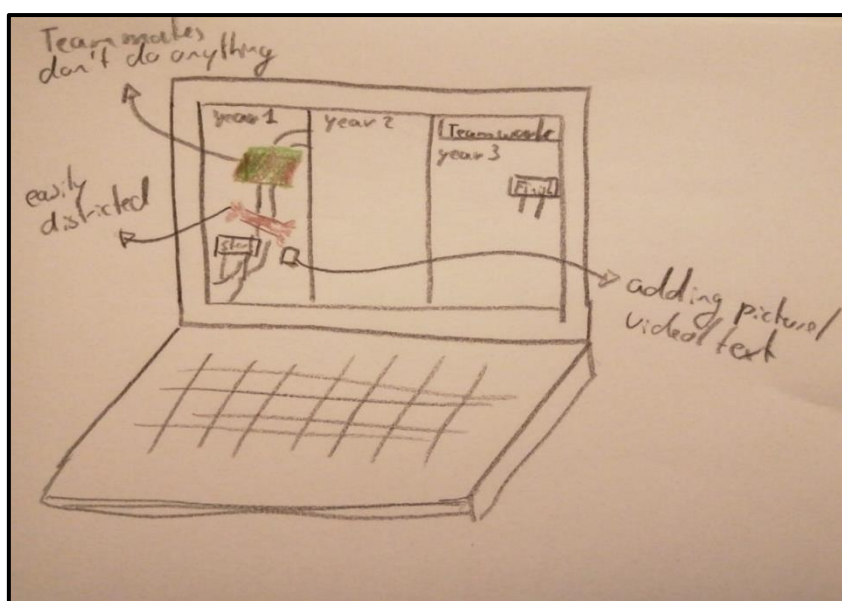


Figure 19: Sketch of design for the Journey Map

## Reflectico

Reflectico is a game that can be played to look back at the past project. By adding a playful element to the reflection process, it is an alternative for writing a reflection report. Through this, it intends to address the problem of students not paying enough attention to their development. The board has different squares have a short assignment on them (Figure 20). Students take turns throwing the dice and moving their pawn. When they land on a square, they need to execute the assignment, writing down their answers on a separate answer sheet.

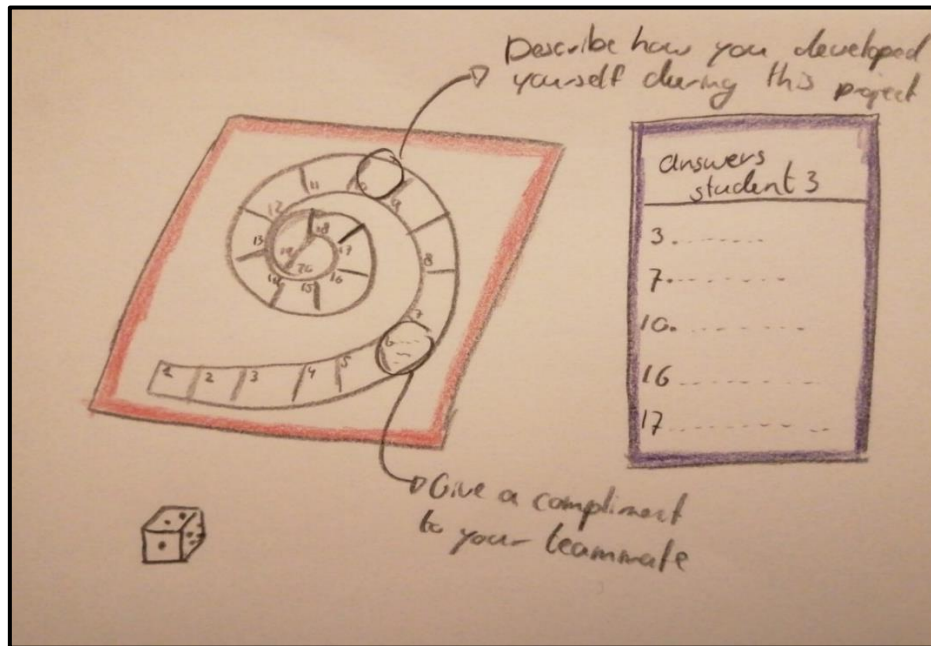


Figure 20: Sketch of design for Reflectico

## Skill Tree

The Skill Tree gives an overview of the different competencies within O&O and how students can work on developing them. Each of these competencies is divided into different components and sorted based on their difficulty, forming a Skill Tree per competency (Figure 21). This provides insight in the pre-set learning goals. To show that they have developed themselves, students can ask teachers to tick of a skill, unlocking higher skills to work on. This allows student to see their progress over time, using the principle of visualising development.

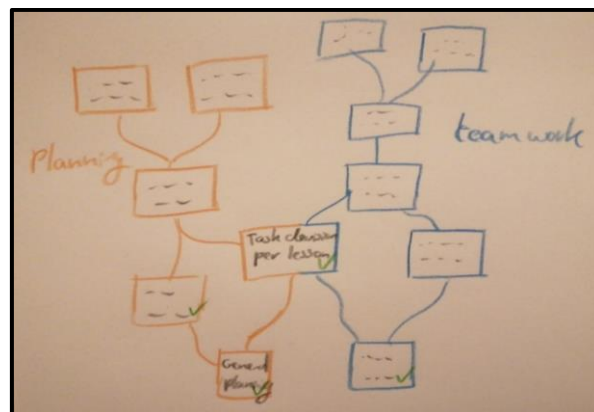


Figure 21: Sketch of design for the Skill Tree

## First evaluation

During the first evaluation session, the six concepts were evaluated with an O&O teacher as outlined in the Methods section. The goal of this session was to come to a selection of concepts to develop further. The advantages, disadvantages and comments that this teacher made about the concepts are discussed below. An overview of this is provided in a table at the end of this section (Table 3). Again, the Dutch quotes have been translated to English, with the original quote provided in the footnotes.

## Challenge Track

When discussing the Challenge Track, the teacher began by mentioning that she currently struggles with grading the process part of the assessment, stating:

*Because I find that the value of the grade in the process really has absolutely no value for me, really nothing at all.<sup>67</sup> (E1)*

Recognizing the necessity of grades within the existing educational system, she noted that the Challenge Track could provide significant support in the grading process. She specifically appreciated how the Challenge Track offered an entirely new perspective on the essence of grades. Which, she indicated, might initially be confusing for students, but she believed that this would actually be something positive. When implementing this concept, the teacher also emphasized that it should be linked to a broader project methodology, allowing students to engage with it at fixed moments.

To add on to this concept, she proposed the idea of allowing students to determine the points they could earn for each Challenge themselves, in a process similar to the planning-poker step in Scrum. Recognizing that this might not always result in exactly ten points, she suggested to make the number of points that can be obtained during a project flexible. To still be able to determine the grade, she suggested to make a separate strip that can be placed next to the point strip, indicating the grade (see Figure 22). She also suggested to make this system dependent on the year the student is in. Potentially creating a point track from one to ten for first-year students and increasing the maximum number of points for higher years.

Points	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Grade	1		2		3		4		5		6		7		8		9		10					

Points	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17			
Grade	1		2		3		4		5		6		7		8		9		10	

Figure 22: An example of two different point tracks with a grade sheet below them to determine the final grade.

To maintain the simplicity of the concept, the teacher recommended keeping the challenges concrete and simple. Therefore, she suggested to just increase the number of challenges in the higher grades instead of making them more complicated.

The teacher also suggested to incorporate additional gamification elements into this concept. Inspired by board games that include a point track, she suggested introducing elements such as bombs or light

<sup>67</sup> Want ik vind die waarde van het cijfer in het proces heeft voor mij echt de totaal geen waarde, echt helemaal niks.

bulbs to either grant extra points or deduce them, making the grading process more playful. However, she acknowledged that if grades directly influence a students' ability to transfer to a next year, such gamification elements might be out of place.

A similar issue arises when students can determine the number of points earned by each of the challenges. In this case, either a high level of trust in the student is required or a lot of checking by the teacher. Letting the teacher check all the assigned points is however a lot of work and it requires a lot of administration. Therefore, she wondered if there would be a way in which this concept can be more self-organized by the students.

Ultimately, this was one of the favourite concepts of the teacher, because she felt that it solves the grading problem that she struggles with.

### Mood Card

Secondly, the Mood Card was discussed. The teacher mentioned that she already considered using brief moments of reflection at the end of every lesson, rather than letting students write a lengthy reflection at the end of the project, as she perceived long reflections to be less valuable for the development of the student. She indicated that the Mood Card would align very well with this approach. Using stickers would, according to her, simplify the reflection process and through that integrate reflection into everyday classroom activities.

However, she did wonder how the use of the Mood Card would lead to a final grade. Different possibilities were explored, such as letting it depend on the number of fields on the Mood Card that a student filled out, the seriousness of the reflection or how positive a student was about his/her performance. She did say that using the Mood Card would help to give insight in the process and therefore simplifies grading an inherently fussy learning process.

A disadvantage mentioned by the teacher was that providing each student with feedback every lesson takes a lot of time. As a solution, the teacher suggested to let the students fill out the Mood Card during every lesson, but that the teacher would only provide feedback during the weekly discussions with the project groups, allowing her to integrate the Mood Card into her routines. By putting the Mood Cards on the table during these discussions, she believed that these would offer an ideal opportunity for students to pause and reflect on their development. Additionally, the use of stickers would make any feedback provided by the teacher visible to the students.

To avoid the use of too many tools, the teacher suggested that the Mood Card replace another tool, such as the log students typically maintain or the end-of-project reflections they write. Especially for first-year students, the teacher said that the Mood Card could effectively replace the log, because students already write a task division at the start of the lesson and writing a log at the end often feels redundant. Furthermore, the Mood Card feels more constructive to the teacher, instead of punishing students for not working hard enough during class, it gives insight in the emotional state of a student. Teachers could review the Mood Cards (of specific students) at the end of the class, using them as a basis for discussions with students during the next class.

Overall, this concept resonated strongly with the teacher and she indicated that she would consider implementing it in her classes.

## Trophy Box

Next, the Trophy Box concept was discussed. From the start the teacher had more trouble seeing how this concept could be implemented in O&O. She pointed out that the Trophy box would mainly serve as a storage solution, without bringing about a structure to help students with reflection. Even as a storage method, the teacher found the idea of individual boxes for each student very cumbersome, leading her to suggest using a single box for each team.

She additionally mentioned that students often struggle with giving each other and themselves compliments, pointing out that it is a skill that first needs to be developed. Furthermore, she raised the issue that sometimes students develop themselves in ways that are hard to pin down and give compliments about.

She did however appreciate the emphasize on capturing positive moments and allowing students to keep track of what was meaningful to them. In cases where students might overlook these moments, the Trophy Box also allows teachers to highlight them with a Compliment Card.

This focus on compliments does however mean that teachers specifically need to remember to write down compliments and make ensure that no student is left out. This means that compliments become less spontaneous, which might diminish their value.

The Trophy Box can additionally be used to show the natural tendency of a student, and show which competencies they predominantly focus on. Such tendencies could be explored in coaching conversations with students. However, the teacher mentioned that this can also this could also be achieved with tokens or markers instead of Compliment Cards.

In general, the teacher did not express much enthusiasm about this concept, as she found it challenging to envision a way in which its implementation would provide significant value to assessment in O&O.

## Journey Map

The Journey Map is a very different kind of tool according to the teacher. It cannot be applied every lesson, but instead takes on the role of a digital portfolio, visualising mapping the process students go through. Consequently, she would prefer to focus on the whole process instead of addressing each competency individually.

One aspect that resonated with the teacher was the use of metaphors to represent various aspects of the process student go through. Given that metaphor thinking might pose challenges to students of that age, she believed that pre-existing objects would serve as a helpful aid, similar to the stickers in the Mood Card. Students can select the object that best shows how they feel about a situation and can add their own explanation of why that object fits best.

Nevertheless, the teacher raised the concern the Journey Map might cause students to primarily focus on external factors influencing the project. They might, for example, complain about disliking the project without considering their own role in it. The teacher therefore suggested to categorize the objects into different types, such as:

- Project related, symbolized by environmental objects;
- Team related, symbolized by character figures;
- Personal attitude related, symbolized by the appearance of an avatar.



This approach encourages students to reflect on all three of these aspects, offering both students and teachers a concrete understanding of the factors influencing project outcomes.

While the teacher expressed appreciation for this solution, she mentioned that it might not be suitable for the lower grades of O&O and she would therefore choose to primarily apply it in the upper grades of O&O.

### Reflectico

The teacher mentioned that a game such as Reflectico can make reflections playful. Nevertheless, she was concerned that the replayability of such a game might be low, fearing that students might get bored with it after having played it once or twice.

To increase replayability, the teacher suggested to use a deck of cards instead of fixed questions. This approach would introduce variability in the questions and questions could be tailored to the specific project. To further enhance the game even more fun, she suggested to incorporate team-building related challenges or dilemma questions.

### Skill Tree

The last of the initial concepts discussed was the Skill Tree. According to the teacher, this concept has the potential to offer students valuable insights into the components that make up various competencies, helping them to further develop in these competencies. In comparison to the 'competentiemonitor', which uses a similar approach, the teacher appreciated how the Skill Tree gives a visual representation.

The teacher did however note that the Skill Tree by itself would be insufficient in letting the students develop their competencies. Furthermore, she expressed concerns regarding the applicability of the Skill Tree to O&O-related competencies. Defining individual components of these competencies could be challenges, because of their fussy nature. Additionally, capturing the development of skills in O&O is complex, given its dependence on a variety of factors, making the improvement of competencies difficult to define.

As an alternative approach, the teacher proposed the idea of having students create their own Skill Trees. We did not discuss this solution in further detail.

### Talent Cards

At the end of the interview, the teacher was asked if she had any additional ideas. The concept she came up with is referred to as Talent Cards from here on onwards. The idea was to focus on the existing talents of the students, instead of things they are not that good at yet. This could encourage students to strengthen their skills. After identifying their talents, students could do challenges that build forth on these. A creative student could for example be asked to come up with 20 different ideas for the next project.

Pre-made Talent Cards could be used, which have a description of the talent on their front and possible challenges on their back. The talent students should be working on during their next project could be determined in discussion with their teacher or by asking their previous teammates what they are good at.

A disadvantage of this concept is however that there is a strong focus on the individual student and their relation to the group becomes less emphasized.

## Overview

In Table 3, the advantages, disadvantages and comments as mentioned by the teacher for each of these concepts are summarised.

Table 3: Overview of the Advantages, Disadvantages and Comments of six concepts from first evaluation session.

		<b>Advantages</b>	<b>Disadvantages</b>	<b>Comments</b>
<b>1</b>	<i>Challenge Track</i>	<ul style="list-style-type: none"> <li>• New perspective on grading</li> </ul>	<ul style="list-style-type: none"> <li>• Grade directly dependant on points students assign</li> <li>• Lot of administration</li> </ul>	<ul style="list-style-type: none"> <li>• Students assigning points to challenges</li> <li>• Keeping challenges simple</li> <li>• Using grade strips to go from points to grade</li> <li>• Use within larger project methodology</li> <li>• Further gamification</li> </ul>
<b>2</b>	<i>Mood card</i>	<ul style="list-style-type: none"> <li>• Integrates reflection in class process</li> <li>• Simplifies reflection</li> <li>• Shows growth</li> <li>• Shows mental state</li> <li>• Easy to check</li> <li>• Makes feedback concrete</li> </ul>	<ul style="list-style-type: none"> <li>• Grading less obvious</li> </ul>	<ul style="list-style-type: none"> <li>• Use as conversation tool</li> <li>• Students use every lesson, teachers sometimes</li> <li>• Different symbols</li> </ul>
<b>3</b>	<i>Trophy Box</i>	<ul style="list-style-type: none"> <li>• Positive moments are emphasised</li> <li>• Shows the natural tendencies of students</li> <li>• Students keep track of meaningful moments</li> <li>• Teachers can point out things students don't see</li> </ul>	<ul style="list-style-type: none"> <li>• Cumbersome if every student has their own box</li> <li>• Students find it difficult to give compliments</li> <li>• It is sometimes hard to give concrete compliments</li> <li>• No structure, just storage</li> <li>• Not spontaneous</li> </ul>	<ul style="list-style-type: none"> <li>• Box per team</li> <li>• Showing what students focus on can be done with for example coins as well</li> </ul>
<b>4</b>	<i>Journey Map</i>	<ul style="list-style-type: none"> <li>• Visualising how they feel about project</li> <li>• Encouraging metaphor thinking</li> <li>• Supporting metaphor thinking with existing images</li> </ul>	<ul style="list-style-type: none"> <li>• Metaphor thinking might be difficult for younger students</li> <li>• Difficult to apply in lower grades</li> <li>• Encourages looking outside themselves</li> </ul>	<ul style="list-style-type: none"> <li>• Not just focussing on one competency</li> <li>• Different categories of images</li> </ul>



		<ul style="list-style-type: none"> <li>Showing concretely which factors influence process</li> </ul>		
5	<i>Reflectico</i>	<ul style="list-style-type: none"> <li>Makes reflecting playfull</li> </ul>	<ul style="list-style-type: none"> <li>Low replayability</li> </ul>	<ul style="list-style-type: none"> <li>Using a deck of cards</li> <li>Adding team-building challenges or dilemmas</li> </ul>
6	<i>Skill Tree</i>	<ul style="list-style-type: none"> <li>Insight in competencies</li> <li>Visualized sub-parts of competencies</li> </ul>	<ul style="list-style-type: none"> <li>Difficult to use by itself</li> <li>No clearly defined skill levels</li> <li>Progress in skills is hard to capture</li> </ul>	<ul style="list-style-type: none"> <li>'Competentie-monitor' uses something similar</li> <li>Alternative: students make their own skill tree</li> </ul>

### Selected concepts

Based on the evaluation interview with the O&O-teacher, the three concepts that the teacher was most positive about were chosen to develop further. These were the Challenge Track, the Journey Map and the Mood Card. The feedback and suggestions of the teacher were used to further improve these concepts. The changes made to each of the concepts will be discussed below.

#### Challenge Track

In Figure 23, the improved concept design for the challenge Track can be found. The design now includes both a point track and a grade strip to convert the number of points into a grade as suggested during the evaluation. The maximum number of points can therefore be different for each project.

The points track is magnetic, so that the (metal) tokens each representing a student of the team will be stuck to the point track and will not get lost. The point track now also starts from 0, so the tokens can already be stuck to the point track at the start of the project. For each completed goal, students also place their token on that goal.

The challenge cards are now also colour coded to show the phase of the project they are relevant to. For simplicity, only group-goals are now used, no longer individual goals. Students can still personalise their journey by determining which of the available goals they want to focus on.

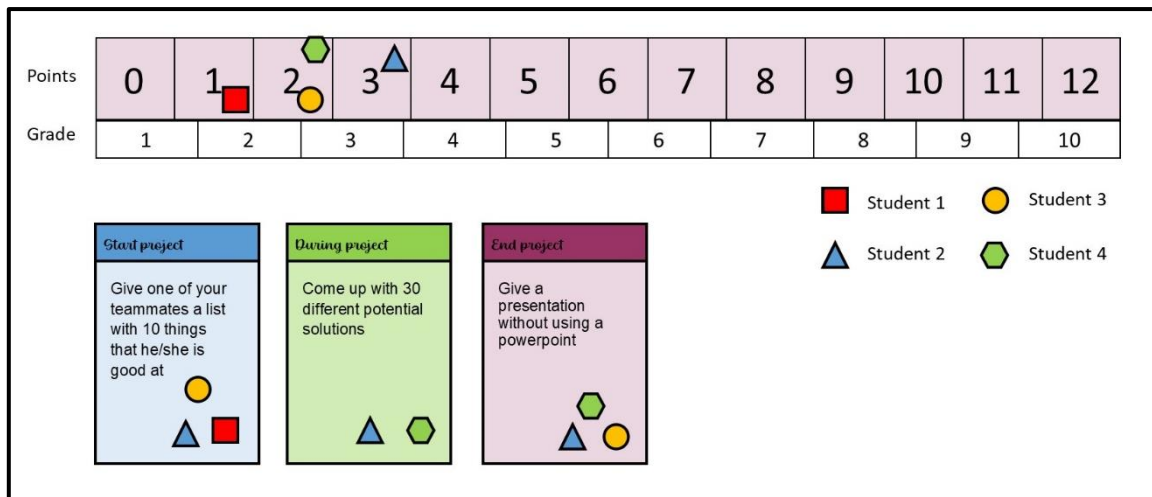


Figure 23: Improved concept design of Challenge Track

The following aspects were not yet decided upon and were therefore discussed during the second evaluation interview:

- How often should new challenge cards be added?
- Should students be allowed to add their own goals to the cards?
- Should new challenge cards be drawn at random?

### Journey Map

To improve the Journey Map, it now focusses on the whole process instead of individual competencies. Furthermore, the objects are now divided into three different categories as suggested by the teacher: environmental, people and emotions. In this way, students can indicate what happened, what role the people around them played and how they felt about this (see Figure 24).

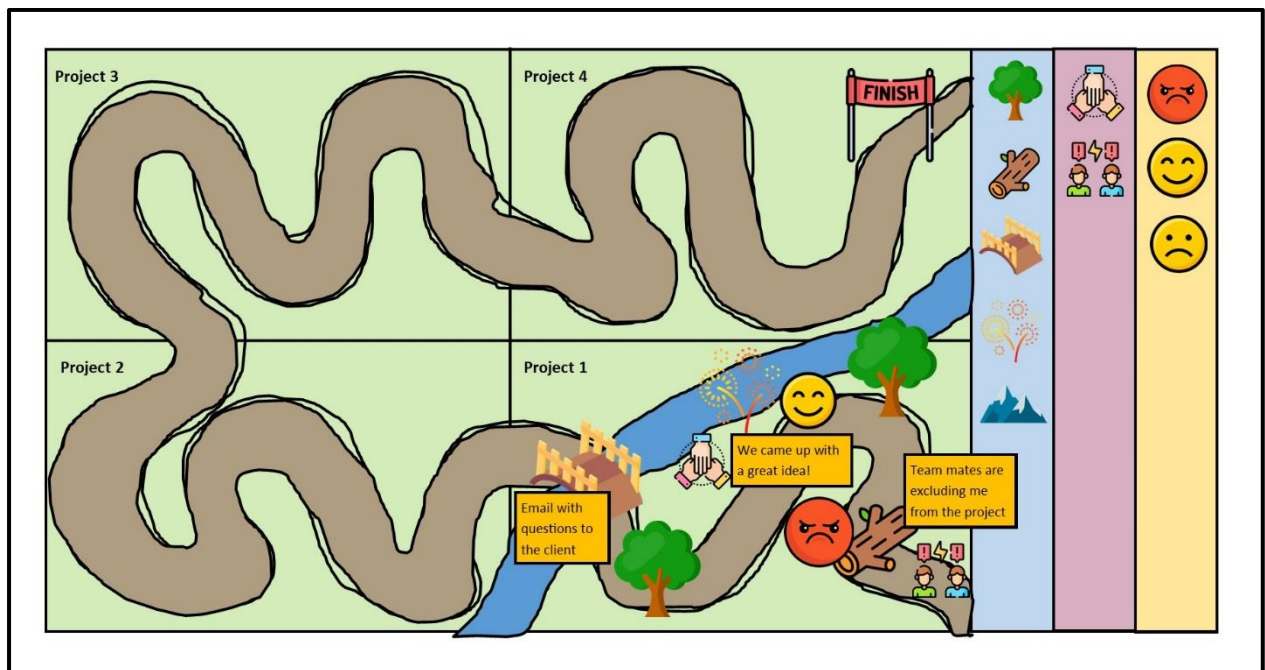


Figure 24: Concept design of journey map images from Flaticon.com are used in the design.

The following question was asked during the second evaluation:

- Should the Journey map be a digital or a physical tool?

### Mood card

To show progression more clearly over the lessons, the Mood Card now is divided into three columns, representing three different moments during the lesson. Students fill in the card at the beginning of the lesson indicating with what mindset they enter the class. If teachers notice that this is a negative one, they can therefore immediately decide to have a conversation with the student. Anything that happens during the lesson can also be documented and at the end of the lesson is another reflection moment. The stickers now have different colours representing the different people: students themselves, peers and teachers (Figure 25).






Lesson	Beginning		Middle		End	
	Sticker	Explanation	Sticker	Explanation	Sticker	Explanation
1		<ul style="list-style-type: none"> <li>• Nice team</li> <li>• Had good sleep</li> </ul>	 	<ul style="list-style-type: none"> <li>• Playing games during class</li> <li>• Angry team-mate</li> </ul>	 	<ul style="list-style-type: none"> <li>• Listened to feedback</li> <li>• Finished introduction</li> </ul>
2						
3						

Figure 25: Stickers in yellow are from the student, blue from a classmate and purple from the teacher.

### Second evaluation

For the second evaluation the three concepts were discussed with another O&O-teacher. After sharing the updated research question and shortly introducing the three concepts, the teacher started off with sharing his thoughts on the research:

*Anyway, it is a big loss, of course, within O&O of what the students get out of it. Especially in the lower grades, it's just like, we come, we deliver something and we go back home. So, it's nice that you now have three possible solutions for that.<sup>68</sup>*  
(E2)

In general, he indicated that the three concepts are filling a gap he experiences. This confirms that the focus on student development is a relevant one.

<sup>68</sup> Sowieso is het een groot gemis, natuurlijk binnen het O&O van het wat de leerlingen eruit halen. Zeker in de onderbouw is het gewoon van ja, we komen d'r, we leveren wat en we gaan weer naar huis.[...] Dus dat is wel fijn dat je daar nu drie mogelijke oplossingen voor hebt.

*The only thing I miss with O&O when it comes to assessment is the process, and all three of them include it. In fact, they are specifically focused on the process.<sup>69</sup> (E2)*

Regardless of what concept would be implemented, the teacher thought that it would be best to start with implementation in the first grade of high school.

*I think that you should just start with these three things in the first year of secondary school because then it becomes a habit.<sup>70</sup> (E2)*

Below, the three different concepts are discussed separately, indicating the advantages, disadvantages, comments and improvements the teacher gave on each of these.

### Challenge Track

The teacher described that the use of the Challenge track can help teachers with grading, which is difficult at the moment according to him:

*How do you assess a project now? Because it's currently somewhat arbitrary. We do know that this group is a six and a half, and this group is a seven. But when parents ask, "why is it a six and a half, why is it a seven?" Sometimes, it feels like we're almost sidelined.<sup>71</sup> (E2)*

Through clearly showing where the grade comes from, the teacher mentioned that the Challenge Track could help with the justification of the grade. Furthermore, he expects that the Challenge Track will bring more focus from the students towards the grade, using the gamification element to push students to work harder from the start of the project. Upon asking if it would not be a problem for students to focus more on the grade, he commented that they are already focused on the grade and does not expect that the Challenge Track will influence the students negatively in that regard. Though he mentioned that a disadvantage of the Challenge Track is that students might miss the point of the challenges and focus so much on getting points instead of focussing on their personal development. Also mentioning that students might focus too much on comparisons amongst themselves. He therefore pointed out that he hopes that at some point the students will realise the value of the challenges outside of the gathering of points.

Next to that, through the challenge cards, teachers can have a significant influence on the projects of students. The teacher suggested to allow the teachers to pick the challenge cards on, for example, a weekly basis. This approach allows teachers to guide students' attention towards specific aspects of their projects without the need for direct criticism. By doing this, the teacher argued that the Challenge Track could be a useful coaching tool and give a lot of freedom to the teacher. To use the Challenge Track in this way, he mentioned that it would be useful if there would be a set of pre-made challenges and an opportunity for teachers to come up with new challenges. According to him, the challenges should stimulate the personal development of the students.

---

<sup>69</sup> Het enige wat ik mis met O&O met beoordelen is het proces, en die zit in alle drie erin. Sterker nog ze zijn juist gefocust op proces.

<sup>70</sup> Ik denk dat je wel met deze drie dingen wel gewoon in de brugklas mee moet beginnen, want dan is het gewoonte.

<sup>71</sup> Hoe ga je nou een project beoordelen? Want dat is nu natuurlijk ook nattevingerwerk. [...] Wij weten wel, dit groepje is een zes en half, dit groepje is een zeven. Maar als ouders vragen: 'waarom is het een zes en half waarom is het een zeven?' Dan staan wij buitenspel bijna, zo voelt het soms.

Lastly, the teacher discussed the number of points per challenge. His opinion was that it would be good that students could get multiple points for each of the challenges. In this way, teachers can differentiate between different levels of quality in completing the challenge and students can also get partial points for unfinished challenges.

In the end, the teacher concluded that he thought this was the concept with the most potential, empathising that there would need to be challenge cards already available, so that teachers do not need to make them all themselves.

### Journey Map

When implementing the Journey Map, the teacher indicates that he would prefer them to be hand-drawn instead of digital. He would also provide a general legend for students to use, but also give them the opportunity to come up with their own symbols, encouraging students to focus on both obstacles and victory's they come across. While he thinks that it's important to let students draw their own journey, he also thought that it would be useful to add the different phases of a research or design project already to the paper. This helps students to increase their awareness of where they are in the project. He would let students add to the Journey map on a weekly basis, starting the week with looking back at the journey map and ending the week with adding to it.

Using the Journey Map, the teacher mentioned that it would be at one glance visible how a student has experienced the project. Because of this, the Journey map can also be shown to the client to give them insight in the process of the students. By showing the experiences the students had during the project, the Journey Map could replace the reflection that students often write at the end of the project, making the process of reflecting more enjoyable for the students. A disadvantage would however be that the teacher expects that it will take some time before the students understand what it expected of them.

An advantage of this tool is that it is personalized, each student can represent their own journey in their own way. However, the teacher mentioned several times that when using a tool like this, students need to be honest and straightforward, something about which he is not sure whether they will be. Part of this is that he emphasized that the Journey Map should reflect the personal Journey of a student, meaning that teachers should watch out for students copying parts of each other's Journey Maps and teachers should not share the Journey Maps of students without their permission.

The teacher thought that the Journey Map could be a useful conversation tool, meaning that he would ask the student questions about it and discuss it during coaching conversations. This also means that he did not think it would be necessary for students to provide written explanations with their drawings. However, the teacher realised that discussing all of the Journey Maps of the students in one class would take up a lot of time.

Because this concept is easily applicable, the teacher indicated that he likes it a lot as well, however because of the significant time investment needed to discuss all the Journey Maps with the students, this was in the end not his preferred concept.

### Mood Card

The Mood Card was the least favourite solution of the teacher. He indicated that it could help with simplification of reflection, however because of this it could become just a formality to students. In other words, students would not take the reflection process seriously.

While the Mood Card can help teachers in determining the process grade and give them insight in the motivation of students throughout the lessons, the teacher said that this concept was less customizable than the other two. Making them more useful for both students and teachers.

### Overview

In Table 4 the advantages, disadvantages and requirements as mentioned by the teacher for each of these concepts are summarised.

Table 4: Overview of the advantages, disadvantages and comments of three concepts from the second evaluation session.

		<b>Advantages</b>	<b>Disadvantages</b>	<b>Requirements</b>
<b>1</b>	<i>Challenge track</i>	<ul style="list-style-type: none"> <li>• Justification grade</li> <li>• Gamification</li> <li>• Motivating to push themselves</li> <li>• Focus attention of students without criticising them</li> <li>• Freedom for teacher</li> <li>• Can be used as coaching tool</li> </ul>	<ul style="list-style-type: none"> <li>• Too much focus on getting points instead of personal development</li> <li>• Comparisons amongst students</li> </ul>	<ul style="list-style-type: none"> <li>• Learning opportunities become more important than points</li> <li>• Pre-made challenge cards</li> <li>• Possibility for teachers to make own challenge cards</li> <li>• Challenges related to personal development</li> <li>• Partial points for low quality or unfinished challenges</li> </ul>
<b>2</b>	<i>Journey map</i>	<ul style="list-style-type: none"> <li>• At one glance visible how a student has experienced a project</li> <li>• Can be shown to client</li> <li>• More enjoyable to draw a reflection instead of writing it</li> <li>• Awareness of where they are within the project</li> <li>• Personalized</li> <li>• Easily applicable</li> </ul>	<ul style="list-style-type: none"> <li>• Having faith in the sincerity of students</li> <li>• Difficult to understand</li> <li>• Copying from other students</li> <li>• Discussing journey map takes a lot of time</li> </ul>	<ul style="list-style-type: none"> <li>• Use on weekly basis</li> <li>• Both obstacles and victories</li> <li>• Reflects a personal journey</li> <li>• Not sharing without permission of student</li> <li>• Conversation tool</li> <li>• No need for written explanation</li> <li>• General legend with symbols</li> <li>• Personal symbols</li> <li>• Physical drawing</li> </ul>

				<ul style="list-style-type: none"> <li>• Phases of research project already drawn</li> </ul>
3	<i>Mood card</i>	<ul style="list-style-type: none"> <li>• Simplifies reflection</li> <li>• Shows motivation throughout lessons</li> <li>• Helps in determining process grade</li> </ul>	<ul style="list-style-type: none"> <li>• Can become a formality</li> <li>• Less customizable</li> </ul>	

### Final design

For the final design it was decided to further develop the Challenge Track. This concept was a favourite of both teachers evaluating the concepts. Furthermore, it solves a concrete problem in O&O: ‘grading of the process’ in a manner that (to my knowledge) has not been done before, giving a new perspective to the meaning of grades. The Challenge Track is furthermore easy to implement and customisable.

### Application

Students start the project with 0 points and the teacher chooses for example two challenges that students can work on. Every week the teacher adds more challenges based on what is happening in the project. In this way the teacher can steer the direction students go into, while still leaving the choice to do so up to the students. When students have (partially) completed a challenge they can go to the teacher to get a stamp on their Stamp Card and they can increase their points on the Points Track. At the end of the project, the points on the Points Track are converted to grades.

### Points Track

Not much has been changed to the Points Track compared to the improved design. The number of points will however be determined by the teacher at the start of the project, so that students can clearly see how far they are on the Points Track.

### Challenge Cards

To make sure that teachers do not need to make all the Challenge Cards themselves, a set of pre-made Challenge Cards will be provided based on common situations in projects. Empty Challenge Cards will be provided as well, so that teachers can come up with their own challenges based on situations they see in the projects.

Each of the Challenge Cards gives a short description of the challenge with below the different levels at which the challenge can be completed (Figure 26). In this way students can also get points for partially finished challenges and students who excelled at a challenge can get additional points. A card number is provided at the bottom of the card to easily refer to the Challenge Card in the Stamp Card.

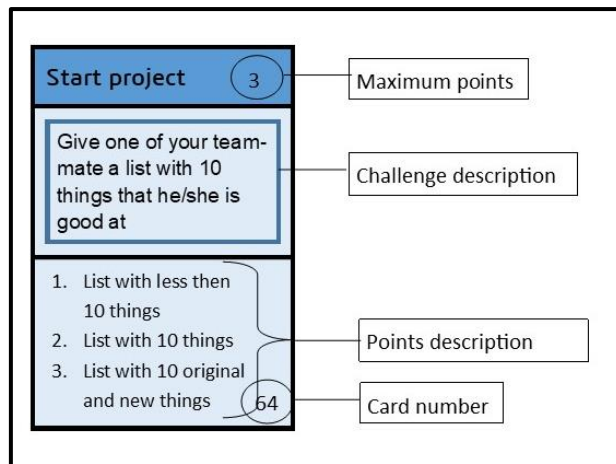


Figure 26: Example of a Challenge Card

### Stamp Card

The Stamp Card can be used by students to indicate what Challenge they completed at what level. As each Challenge Card has its own number, students can fill out that number on their Stamp Card. Teachers can give students stamps when they can (partially) completed a challenge and students can keep track of the points they collected on the Point Track (Figure 27). The Stamp Card is individual for each student.









Name:			
Team:			
Challenge	Level		
	1	2	3
21			
13			
7			
64			
5			

Figure 27: Example of a Stamp Card

### Learning line

Because of the flexibility of the Challenge Track, it can be applied differently in different grades of O&O. In the lower grades of O&O, challenges could for example be mostly determined and picked by the teacher to stimulate students to develop themselves in certain directions. However, in the higher grades of O&O, students could also come up with their own challenges, encouraging them to think about their own personal development and set SMART goals for themselves. Because students in higher grades are more mature and can handle more responsibilities (see interviews), then determining their own goals in discussion with their teacher will be less likely to result in problems of students deliberately setting low goals to get a high grade.

## Discussion & Conclusion

This study aimed to develop a tool to support teachers in providing high-quality assessment in O&O. Findings from interviews and the focus group revealed two critical needs: (1) the necessity for clarifying learning goals and (2) the use of assessment to support student development. The latter was used as input for the design process, resulting in the development of the Challenge Track. This tool is grounded in the principle of ‘creating space for growth’ to support the development of students.

One of the key contributions of this tool is its potential to support teachers in grading the process assessment, while simultaneously clarifying the assessment process for all involved stakeholders. Though this the Challenge Track influenced student development and quality assurance as well as playing a role in motivating students.

The main intention in the design process was to develop a tool that can support teachers in the facilitation and enhancement of student development. The Challenge Track intends to do so by providing students with concrete challenges that are assumed to trigger their development. This reminds of the use of gamification elements, which can be used to increase motivation and stimulate learning (Eltahir et al., 2021). It remains however unclear what the effect of these kinds of challenges would be on student development. This would be an interesting and relevant avenue for future research.

Beyond its focus on student development, the Challenge Track introduces a novel strategy for quality assurance as well. Participants in both evaluation sessions expressed difficulties with grading the process component of the assessment, a challenge unaddressed in the interviews or focus group. However, issues with time investment and clarity of learning goals were addressed here. This is in line with the findings of Chan and Luk (2022), indicating that the assessment of personal and professional development competencies is different and maybe even impossible due to the abstract nature of these competencies. An alternative approach suggested in the Reframing chapter, derived from interview and focus group findings, was to emphasize effort-based standards over the measurement of performance. This is precisely the approach the Challenge Track adopts, allowing for flexibility in student development outcomes while concentrating assessment on the effort invested. The Challenge Track quantifies this effort through the number of challenges completed by each student.

Lastly, the Challenge Track is designed to motivate students through the use of grades as extrinsic rewards. In the second evaluation session the teacher expected a stronger focus on the grade when using the Challenge Track, viewing this as mostly positive. However, he also raised concerns about a too strong focus on collecting points. The interviews and focus group indeed revealed that students might become dishonest or stop thinking for themselves because of the use of grades. This would potentially diminish the value of the Challenge Track. Similar concerns have been raised in the literature, discussed as the negative effects of washback (Baird et al., 2017). This argues for the careful design of the challenges, making sure that the learning goals of O&O align with encouraged behaviour of students (Baird et al., 2017; Biggs, 2003). The design principles for these challenges could be another avenue for future research.

Figure 28 shows in red the three pathways the Challenge Track intends to use, improve construct mastery through a focus on student development, checking the effort a student exerted for quality assurance purposes and using motivational influences to increase a student’s effort.

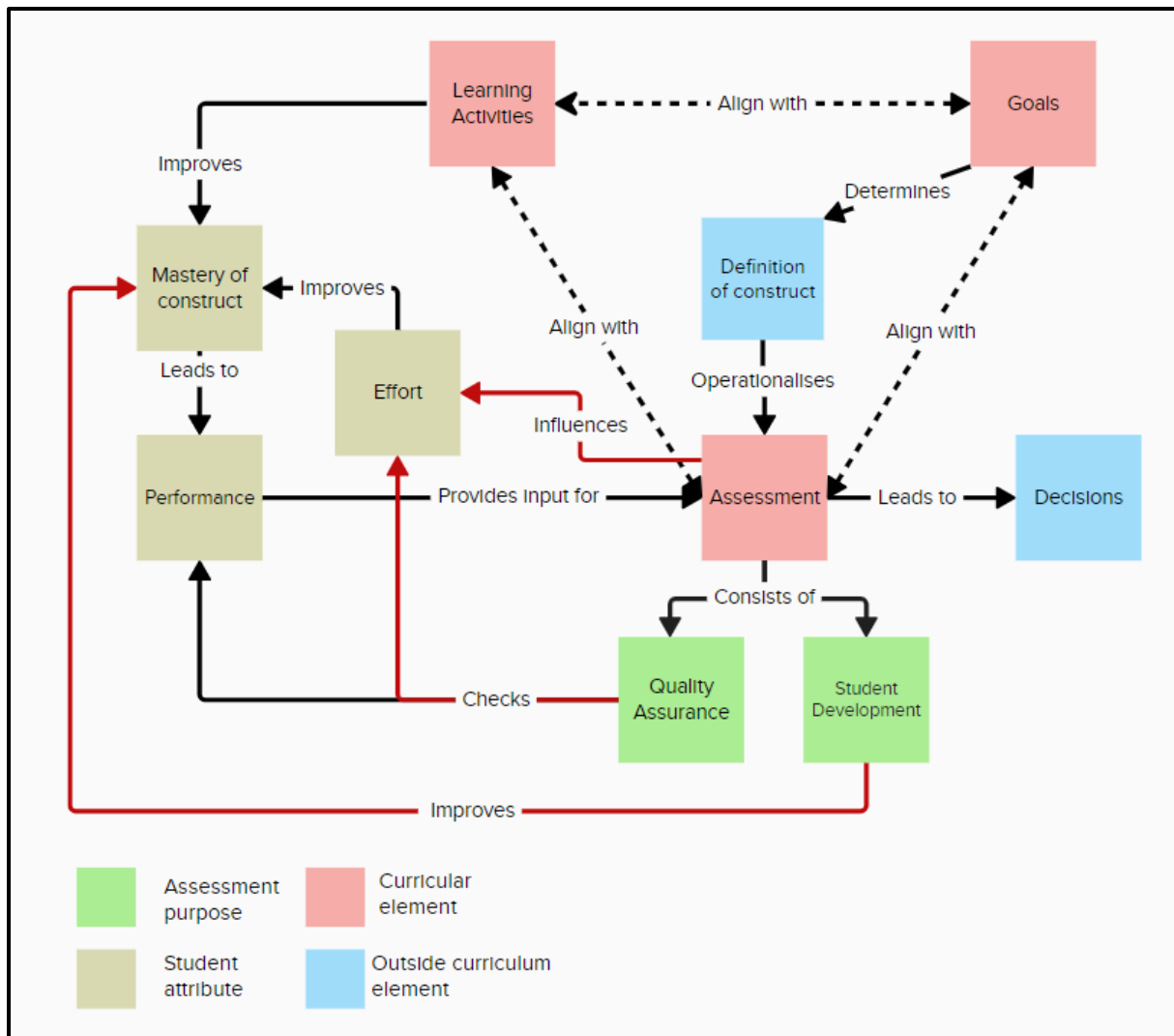


Figure 28: Pathways the Challenge Track intends to use are indicated in red.

## Limitations

While contributing valuable insights about assessment in O&O, this research is not without limitations. There needs to be carefully considered in interpreting the findings and drawing conclusions.

A primary limitation of this study arises from the very nature of thematic analysis, employed in this study. This method relies heavily on the interpretation of the researcher, introducing subjectivity in the analysis of the results. While efforts were made, as described in the Methods, to faithfully represent the viewpoints of the participants, it is important to recognize the influence of the researcher's perspective of the identification and interpretation of the themes. Consequently, different researchers with different viewpoints might have arrived at different conclusions.

However, similar to the assessment of students, much valuable information gets lost when attempts are made to present a completely objective view (Schuwirth & Van der Vleuten, 2012). While my experience as an O&O teacher might have influenced my interpretations of the results, it also provided me with the necessary background to place the findings in a relevant context. This aided me in providing a holistic perspective on assessment in O&O.

As Schuwirth and Van der Vleuten (2012) also argued with regards to subjective assessments, it would be valuable to provide multiple perspectives on this subjective assessment. It would therefore be relevant to for future research to be compared and contrasted with findings from this research.

Another significant constraint is the relatively limited sample size of 22 participants. This sample size restricts the generalisability of the findings, as they may not be representative of assessment at other O&O schools. It would therefore be recommended to conduct similar research with a larger sample of participants to see if the results are more broadly applicable. It would furthermore be interesting to introduce a larger variety of participants, such as Technasium foundation staff, school management, technical assistants (TOA), clients and stakeholders from other high school CBL courses.

Additionally, the research faced a homogeneity issue with the participant pool. Students from only two schools and teachers from only one school were involved during the interviews. This shared school affiliation could potentially introduce biases or specific contextual factors that do not generalize well. The research sought to address this issue through the introduction of a more diverse participant pool in the focus group and evaluation sessions. Nevertheless, this could have significantly influenced the results of this research.

### Recommendations

The blueprint for the Challenge Track outlined in the Design chapter of this report can be used to further develop and implement the ideas presented in this research. This section lays out the recommended steps towards implementation of the Challenge Track in O&O education.

It is recommended that the design of the Challenge Track, as outlined in the Design chapter, undergoes further evaluation in collaboration with stakeholders. This should include the active involvement of students as the second main user of the Challenge Track (Steen, 2011). To make the design tangible for the participants, the creation of a low-fidelity prototype is advised. The deliberate choice to create a low-fidelity prototype serves the purpose of conveying the unfinished and evolving nature of the concept to the participants. This approach fosters an environment where participants are not only comfortable but also encouraged to propose major changes to the design. As part of this prototype, it is recommended to provide an exemplary set of about 15 challenges cards with a variety of challenges, this can give participants some insight into potential challenges.

After implementing the feedback from these evaluation sessions, it is recommended to run a first pilot with the Challenge Track. To ensure the challenges align with the needs and preferences of students and teachers, it is strongly recommended to involve them in the process of crafting these challenges. In this way, the needs of students and teachers can be properly reflected in the challenges.

For this initial pilot it would be specifically relevant to experiment with and focus on (1) the number of points on each of the challenge cards (2) the total number of points over the project and (3) the use of personal, team or class challenges. The outcomes of this pilot should be reviewed in collaboration with the participating teachers and students to refine and optimize the Challenge Track's design.

Building upon the insights gained from the first pilot, a second pilot implementation is recommended. This phase should prioritize an exploration of the potential benefits associated with the Challenge Track, with a specific focus on areas such as: (1) Increases in motivation, (2) increases in assessment clarity and (3) increases in student development. The findings from this second pilot should once again be assessed in conjunction with the engaged students and teachers to further fine-tune the design.

This systematic and collaborative approach, involving multiple phases of evaluation and piloting, ensures that the Challenge Track is refined and optimized, aligning it more closely with the requirements and preferences of the stakeholders and increasing its effectiveness in O&O education.

## References

- AERA, APA, & NCME. (2014). *Standards for Educational and Psychological Testing: National Council on Measurement in Education*.
- Amberscript. (n.d.). *Amberscript: Audio & Video Transcription | Speech-to-text*. Retrieved 30 September 2023, from <https://www.amberscript.com/en/>
- ATLAS.ti. (n.d.). *ATLAS.ti | The #1 Software for Qualitative Data Analysis*. Retrieved 30 September 2023, from <https://atlasti.com/>
- Baird, J. A., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment and learning: fields apart? *Assessment in Education: Principles, Policy and Practice*, 24(3), 317–350. <https://doi.org/10.1080/0969594X.2017.1319337>
- Biesta, G. (2022). *Wereldgericht onderwijs* (2nd ed.). Phronese.
- Biggs, J. (2003). Aligning teaching and assessing to course objectives. *Teaching and Learning in Higher Education: New Trends and Innovations*, April, 13–17.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Blume-Bos, A., van der Veen, J. T., & Boerman, P. L. J. (2020). Engineering in Dutch schools: Does it effect study choice? *SEFI 48th Annual Conference Engaging Engineering Education, Proceedings*, 84–92.
- Boor, I., Gerritsen, D., De Greef, L., & Rodermans, J. (2021). *Meaningful Assessment in Interdisciplinary Education: a practical handbook for university teachers* (first). Amsterdam University Press B.V. <https://doi.org/10.5117/9789463729048>
- Braun, V., & Clarke, V. (2006). Qualitative Research in Psychology Using thematic analysis in psychology Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <http://www.tandfonline.com/action/journalInformation?journalCode=uqrp20> <http://www.tandfonline.com/action/journalInformation?journalCode=uqrp20>
- Broadfoot, P., & Black, P. (2004). Redefining assessment? The first ten years of assessment in education. *Assessment in Education: Principles, Policy and Practice*, 11(1), 7–26. <https://doi.org/10.1080/0969594042000208976>
- Chan, C. K. Y., & Luk, L. Y. Y. (2022). Going ‘grade-free’?—Teachers’ and students’ perceived value and grading preferences for holistic competency assessment. *Higher Education Research and Development*, 41(3), 647–664. <https://doi.org/10.1080/07294360.2021.1877628>
- De Vijlder, F., Bakker, D., & Van den Blink, M. (2014). *Innoveren vanachter de keukentafel: Een onderzoek naar de ontwikkeling van het technasium 2003-2013*. Convoy Uitgevers B.V.
- Eltahir, M. E., Alsalhi, N. R., Al-Qatawneh, S., AlQudah, H. A., & Jaradat, M. (2021). The impact of game-based learning (GBL) on students’ motivation, engagement and academic performance on an Arabic language grammar course in higher education. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-020-10396-w>

- Folmer, E., Koopmans-van Noorel, A., & Kuiper, W. (2017). *Curriculumspiegel 2017*. 401. [www.slo.nl](http://www.slo.nl)
- Gallagher, S. E., & Savage, T. (2023). Challenge-based learning in higher education: an exploratory literature review. *Teaching in Higher Education*, 28(6), 1135–1157. <https://doi.org/10.1080/13562517.2020.1863354>
- Gibb, S. (2014). Soft skills assessment: theory development and the research agenda. *International Journal of Lifelong Education*, 33(4), 455–471. <https://doi.org/10.1080/02601370.2013.867546>
- Green, A. (2007). Washback to learning outcomes: A comparative study of IELTS preparation and university pre-session language courses. *Assessment in Education: Principles, Policy and Practice*, 14(1), 75–97. <https://doi.org/10.1080/09695940701272880>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Hodges, B. (2013). Assessment in the post-psychometric era: Learning to love the subjective and collective. *Medical Teacher*, 35(7), 564–568. <https://doi.org/10.3109/0142159X.2013.789134>
- Hughes, C., & Barrie, S. (2010). Influences on the assessment of graduate attributes in higher education. *Assessment and Evaluation in Higher Education*, 35(3), 325–334. <https://doi.org/10.1080/02602930903221485>
- Lechner, J. (2012). *Coachen op het technasium: een dubbel belang*.
- Long, H., & Wang, J. (2022). Dissecting Reliability and Validity Evidence of Subjective Creativity Assessment: A Literature Review. In *Educational Psychology Review* (Vol. 34, Issue 3). Springer US. <https://doi.org/10.1007/s10648-022-09679-0>
- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy and Practice*, 14(2), 149–170. <https://doi.org/10.1080/09695940701478321>
- NWV. (n.d.). *Actualisatie examenprogramma's*. Retrieved 28 November 2022, from <https://www.actualisatienwv.nl/proces/examenprogrammas>
- Paulus, P. B., & Brown, V. R. (2003). Enhancing Ideational Creativity in Groups: Lessons from Research on Brainstorming. In P. B. Paulus & B. A. Nijstad (Eds.), *Group Creativity: Innovation Through Collaboration* (pp. 110–136). <https://doi.org/10.1093/acprof:oso/9780195147308.003.0006>
- Pellegrino, J. W., & Chudowsky, N. (2003). FOCUS ARTICLE: The Foundations of Assessment. *Measurement: Interdisciplinary Research & Perspective*, 1(2), 103–148. [https://doi.org/10.1207/s15366359mea0102\\_01](https://doi.org/10.1207/s15366359mea0102_01)
- Prins, G. T., Vos, M. a J., & Pilot, A. (2011). Leerlingpercepties van onderzoek & ontwerpen in het technasium. In *FISME Scientific Library* (Issue FISME Scientific Library 69). <http://www.fisme.science.uu.nl/toepassingen/20069>
- Schalk, H., & Bruning, L. (2014). *Handreiking schoolexamen Onderzoek & ontwerpen havo / vwo*.
- Schuwirth, L. W. T., & Van der Vleuten, C. P. M. (2012). General overview of the theories

- used in assessment: AMEE Guide No. 57. In *Medical Teacher* (Vol. 33, Issue 10). <https://doi.org/10.3109/0142159X.2011.611022>
- Sluismans, D. (2020). Toetsing als kans voor leren. *NRO Kennis Rotonde*, 1–23.
- Steen, M. (2011). Tensions in human-centred design. *CoDesign*, 7(1), 45–60. <https://doi.org/10.1080/15710882.2011.563314>
- Stichting Technasium. (n.d.). *Technasium in cijfers*. Retrieved 18 August 2021, from <https://www.technasium.nl/technasium/technasium-in-cijfers/>
- Stichting Technasium. (2021). *Onderwijsfilosofie van het Technasium op hoofdlijnen*.
- van den Akker, J. (2003). Curriculum Perspectives: An Introduction. In J. van den Akker, W. Kuiper, & U. Hameyer (Eds.), *Curriculum Landscapes and Trends*. Springer-Science+Business Media, B.V. <https://doi.org/10.1007/978-94-017-1205-7>
- Van der Vleuten, C. P. M., Norman, G. R., & De Graaff, E. (1991). Pitfalls in the pursuit of objectivity: issues of reliability. *Medical Education*, 25(2), 110–118. <https://doi.org/10.1111/j.1365-2923.1991.tb00036.x>
- Visser, T. (2023). *Procesbeoordeling bij O&O; Objectief en eerlijk?* University of Twente.
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3–14. <https://doi.org/10.1016/j.stueduc.2011.03.001>
- Wiliam, D. (2017a). Assessment and learning: some reflections. *Assessment in Education: Principles, Policy and Practice*, 24(3), 394–403. <https://doi.org/10.1080/0969594X.2017.1318108>
- Wiliam, D. (2017b). Learning and assessment: a long and winding road? *Assessment in Education: Principles, Policy and Practice*, 24(3), 309–316. <https://doi.org/10.1080/0969594X.2017.1338520>
- Wols, I. (2022). *Kennisuitwisseling binnen het Technasium*. University of Twente.



## Appendix

### Appendix A: Overview of Quality Criteria

PURPOSE	CRITERIA	EXPLANATION	SOURCE
<b>QUALITY</b>	Quality of results	Students are assessed on the quality of their final result, such as source utilization, adherence to client specifications, and quality of their presentation	S1, S2, S4, S5, S7, S8, T1, T2, U3, FG-SA
	Assessment of development	The growth and development of students is assessed	S1, S2, S3, S4, S5, S6, T1, T2, U1, U2, U3, FG-SA
	Assessing effort and behaviour	How much time and effort a student put into the project should be reflected in their final assessment	S1, S2, S3, S4, S5, S7, S8, S9, S10, T1
	Holistic perspective	The teacher is not too concerned with the details of the result, but instead mainly focuses on the depth of a project's content	S3, S8, U3
	Clear assessment criteria	It is clear to students what they are assessed on and why they get the assessment that they get	S2, S3, S4, S5, T1, T2, U3
	Authority teacher	The teacher is ultimately responsible for the assessment of the students and therefore has the final say	S7, S8, T1
<b>MOTIVATION</b>	Building self-esteem	Students should be able to be proud of themselves and have faith that they can approach other problems as well	T1, U3, FG-RA
	Pushing students	Students are motivated by the assessment to work harder	S1, S2, S3, T1, U1, U2, FG-SA
	INTRINSIC MOTIVATION	STUDENTS ARE MOTIVATED BY FEELING THAT O&O IS PERSONALLY RELEVANT TO THEM	T2, FG-RA, FG-PS
<b>STUDENT DEVELOPMENT</b>	Preparation	O&O prepares students for things they might face later on	S1, FG-SA
	Connecting personal learning goals to	By themselves, the competencies as described by Technasium are not always	T2, U1, U3

Technasium competencies	relevant, therefore they should be connected to personal goals	
Personalized learning goals	Students can work on their own personal learning goals	T2, U1, U3, FG-RQ
Valuable feedback	Students are given valuable feedback at the time when they need it to develop themselves.	S5, S6, T1, T2, U1, U2, U3, FG-SA
In mutual discussion with Student	Decisions on assessment are made in discussion with students, they get a say in the assessment process	S1, S2, S3, S4, S6, S7, T1, T2, U1, U2, U3, FG-RA
Dependant on age and maturity of students	The assessment practices depend on the age and maturity of a student.	S3, T1, T2
Focus on Talent	Teachers focus on the strengths of the student	FG-IG

## Appendix B: Overview of Obstacles

PURPOSE	PROBLEM	EXPLANATION	SOURCE
<b>QUALITY</b>	Unclear assessment criteria	It is unclear to the parties involved what the assessment criteria for O&O are.	S1, S4, S6, S9, S10, U1
	Personal development is not tangible	You cannot predict beforehand how students will develop themselves and what obstacles they will run into	S3, T2, U2
	Unfair	Students feel that the assessment is unfair.	S1, S4, S6, S7, S8, S9, S10
	Unawareness Teachers	The teacher does not know how hard the students work and what they are doing during class time	S1, S2, S3, S4, S5, S7, T1, U1, U3
<b>MOTIVATION</b>	No thinking	Students are collecting points or act in a way that they think the teacher wants them to instead of thinking for themselves	S3, T1, T2, U3
	Dishonesty	Students are dishonest about the work they put in, in the hope of getting a higher grade	S1, S2, S3, S8, S9, T1
<b>STUDENT DEVELOPMENT</b>	Insufficient feedback	Student do not know how they can improve themselves because of the lack of feedback.	S3, S5, S9, S10
	Good assessment takes a lot of time	It takes a lot of time for teachers to assess the work of students. This mean that students sometimes get their assessment very late, at which point it is less useful.	S2, S3, S6, S9, S10, T1, T2, U1, U2, U3, FG-RA, FG-RQ
	Students don't find the process as important	Students find it more interesting to focus on making a product then developing themselves.	T2
<b>QUALITY – STUDENT DEVELOPMENT</b>	Teacher is also responsible for results	The choice of projects, coaching of students and feedback of the teacher can influence the quality of the results of the students.	

## Appendix C: Overview of Solutions

PURPOSE	SOLUTION	EXPLANATION	SOURCE
<b>QUALITY</b>	Teachers' logbook	Teachers keep a logbook of what students are working on during the class hours to help them give a fairer assessment	S1, S4, S5, S7, S8

	Noting down task division	The students or teachers keep track of the planned and actual task division	S1, S2, S7
	Making the rubric with the students	The rubric is developed together with the students, so they are very much aware of the assessment criteria	U1, U2, U3
<b>MOTIVATION</b>	No points on the rubric	Because there are no points of the rubric, students are less inclined to just do what the rubric says	S3, T2
	Accountability for personal development	Students should feel responsibility for their own personal development	S3, T1, T2, U1, U3, FG-RQ
<b>STUDENT DEVELOPMENT</b>	Flexible rubric	Certain parts of the rubric are fixed, but certain parts can be chosen by the student	T2, U1
	Intermediate assessments	Have small 'assessment' conversations with the teacher during or at the end of some lessons	S2, S4, S5, S7, S8, S9, S10
	Comparative judgement	The work of students is compared to make it easier to assess it.	U1, U2
	Older O&O students help with coaching	O&O students from higher grades help teachers with the coaching of younger O&O students, so that there is more time for assessment	S3
	Ticking off skills	Students need to show their mastery of individual skills throughout their learning trajectory.	U3
	Self-assessment	Students use a tool to help them assess themselves	FG-IG
	Learning line	Documents outlining the learning goals for each year/project	FG-IG
	Portfolio	Students keep a portfolio to track their personal and professional development	T2, U3, FG-IG
	Explanations	Teachers explain to students how certain aspects of O&O work	S1, S2, S3
<b>PERSONAL DEVELOPMENT - MOTIVATION</b>	Personal development separate from assessment	The personal development of students is expected, but not taken into account in the assessment	
<b>QUALITY – STUDENT DEVELOPMENT</b>	Assessment of teachers	Students assess their teachers as well and give them feedback	