



IARTEM *e-Journal* Volume 8 No 1/2

Volume 8 No 1/2

Systematic, digital student feedback for differentiated teaching

Stefan Ting Graf &
Dorthe Carlsen

Abstract

The article reports results from a qualitative study of Elevbaro, a prototype of a digital tool for student feedback developed in connection with the demonstration school project, Inclusion, and differentiated teaching in digital learning environments. At the same time the study represents the first step of validating Elevbaro as a systematic feedback tool. There is general consensus that feedback is central to the quality of teaching, but the focus on and the exploration of systematic student feedback is an overlooked topic. Especially as regards differentiated teaching and complex teaching patterns, there is a need for supplementary and digital monitoring of a group of students and of individual students. The article examines how students and teachers understand and use Elevbaro, which is built on frequent ratings of five set statements in connection with teaching over a certain period of time. There is reason to assume that the usage of Elevbaro affects teachers' professional assessment of students as well as students' reflections of their work and acquisition process. This requires an active approach to Elevbaro and the integration of the tool into the teacher's usual assessment and teaching practice. In our view, there is didactic potential for teachers' data-informed teaching development.

Keywords: student feedback, digital tool, differentiation, teaching, learning analytics.

Resumo

O artigo relata resultados de um estudo qualitativo de Elevbaro, um protótipo de uma ferramenta digital que fornece *feedback* aos alunos, desenvolvida em conexão com o projeto de demonstração da escola, Inclusão e ensino diferenciado em ambientes de aprendizagem digital. Ao mesmo tempo, o estudo representa o primeiro passo para validar o Elevbaro como uma ferramenta de *feedback* sistemática. Há consenso geral de que o *feedback* é fundamental para a qualidade do ensino, mas o foco no *feedback* sistemático dos alunos e a exploração do mesmo é um tópico negligenciado. Especialmente no que se refere ao ensino diferenciado e aos padrões complexos de ensino, há necessidade de monitoramento suplementar e digital de um grupo de alunos e de alunos individualmente. O artigo examina como os alunos e os professores entendem e utilizam o Elevbaro, que é construído com base em avaliações frequentes de cinco afirmações definidas relacionadas ao ensino, durante um determinado período de tempo. Há razão para supor que o uso de Elevbaro afeta a avaliação profissional dos professores pelos alunos, bem como as reflexões dos alunos sobre seu trabalho e seu processo de aquisição. Isso requer uma abordagem ativa do Elevbaro e a integração da ferramenta na prática usual de avaliação e ensino do professor. Em nosso ponto de vista, existe um potencial didático para o desenvolvimento do ensino baseado em dados, pelos professores.

Palavras-chave: *Feedback* do aluno, ferramenta digital, diferenciação, ensino, learning analytics.

Resumen

El artículo presenta resultados de un estudio cualitativo de Elevbaro, un prototipo de una herramienta digital que proporciona *feedback* a los alumnos, desarrollada en conexión con el proyecto de demostración de la escuela, Inclusión y enseñanza diferenciada en ambientes de aprendizaje digital. Al mismo tiempo, el estudio representa el primer paso para validar el Elevbaro como una herramienta de *feedback* sistemático. Hay consenso general de que el *feedback* es fundamental para la calidad de la enseñanza, pero el foco en el *feedback* sistemático de los alumnos y la exploración del mismo es un tópico olvidado. Especialmente en lo que respecta a la enseñanza diferenciada ya los patrones de enseñanza complejos, es necesario un seguimiento complementario y digital de un grupo de estudiantes y de estudiantes individuales. El artículo examina como los alumnos y los profesores entienden y utilizan el Elevbaro, que es construido con base en evaluaciones frecuentes de cinco afirmaciones definidas relacionadas a la enseñanza, durante un determinado período de tiempo. Hay razones para suponer que el uso de Elevbaro afecta a la evaluación profesional de los profesores por los alumnos, bien como reflexiones de los alumnos sobre su trabajo y el proceso de adquisición. Esto requiere un abordaje activo de Elevbaro y la integración de la

herramienta en la práctica usual de evaluación y enseñanza del profesor. En nuestra opinión, existe un potencial didáctico para el desarrollo de la enseñanza apoyada en los datos.

Palabras clave: Feedback a los alumnos, herramienta digital, enseñanza diferenciada, learning analytics.

1. Introduction

While teachers' feedback to students is greatly in focus, feedback *from* students to teachers holds a less prominent position in the educational discourse. In the so-called demonstration school project, *Inclusion and differentiated teaching in digital learning environments* (IDDL), which was carried out between 2013 and 2015 along with similar projects within the framework of the national digitalization strategy in Denmark, we have developed a prototype of a digital tool which provides teachers with a data-informed basis for differentiating their teaching through a systematic compilation of student feedback. The tool is called Elevbaro and can be accessed at www.elevbaro.dk. The development of the tool is partly inspired by – and differs from – other similar tools, and is partly based on assessment and general feedback assumptions, as well as on the concept of inclusion and differentiated teaching within the project. Elevbaro was a sub-element in a comprehensive multidimensional intervention of four phases over three semesters in five schools. To support teachers, the intervention homepage describes and motivates Elevbaro theoretically and practically (auuc.demonstrationsskoler.dk/uvd). Intervention consultants have introduced the tool and followed up on its usage within the scope of the overall intervention in roughly one school year (cf. Graf, 2016).

This article presents results from an explorative case study based on observations and interview whose overall purpose is to study teachers' and students' use and understanding of the tool on one hand and to assess and ensure the quality of the tool on the other hand. First, we argue why there is a need for student feedback based on results of our literature search, on the need for systematic student monitoring in relation to differentiated teaching and on the need for digital remote monitoring especially in project-oriented teaching. Second, we describe and justify the design of the tool and the underlying assumptions. Third, we present results from our qualitative research into how teachers and students use and understand Elevbaro. Moreover, it is important for the teacher's perspective to examine how the tool is integrated into the teaching, how the teacher interprets Elevbaro data and how to act upon it. Thus, we continually discuss the didactic possibilities and limitations of the tool from a general feedback and assessment perspective.

1.1. Student feedback in a didactic¹ perspective is under-examined

It is generally accepted today that teachers' feedback to students is one of the single most important factors when it comes to improving teaching and learner outcome (cf. Hattie & Timperley, 2013; Barber & Mourshed, 2007). Students' feedback to teachers, however, especially feedback on their participation in teaching activities, is still under-examined. Searches on Bibliotek.dk, Forskningsbasen.dk, Bybysys (Oria.no) and ERIC using *pupils' or students' feedback* and similar key words relating to primary and lower secondary school yield only few peer-reviewed articles. If the search is broadened, a number of research fields can be identified internationally which integrate student feedback. One branch that carries out research into pupils' or students' feedback deals with studies of quality standards of institutions, including satisfaction assessments. Another research branch is carrying out school effectiveness studies through ratings from pupils or students. A third branch is examining connections between student feedback and self-assessment in relation to learning outcomes. A new and fourth research branch is data-informed development and the identification of student feedback as a source of data (learning analytics). This branch, too, however, is primarily concerned with student performance (Robinson et. al. 2009). What these approaches have in common is that they chiefly focus on outcomes and subscribe to sociological and organizational theory, and their results primarily meet educational policy demands regarding the design and management of effective and accountable educational systems. These approaches have limited value when it comes to the concrete, didactic challenges facing teachers. Therefore Elevbaro is meant to be a tool for the teacher and his or her students as individuals and as a group in order to enhance and develop differentiated teaching and learning.

1.2. Need for monitoring through student feedback

The quality of teaching is closely connected to how successfully the teaching is differentiated (Egelund, 2010). Our understanding of inclusion as academic and social participation in teaching is highly dependent on the extent to which the teaching is successfully differentiated. The concept of inclusion in the IDDL project is described in *Fire differentieringspositioner – teoretisk og empirisk set* (Carlsen & Graf, under publishing). One of the challenges when it comes to differentiation is how the teacher will continually be able to gain knowledge about which students are experiencing what kind of challenges. It seems helpful to distinguish between the teachers' *professional image of a student* and their *situative judgment* of a student. Teachers' images of students' academic and social participation are formed as more or less stable views of students which shape both new estimations of students as well as expectations of them. This general image is based partly on a number of situative judgments in concrete teacher-student interactions, and partly on teachers' access to various data

¹ The expressions didactics and didactic are used in the European continental sense.

such as grades/marks, test results, satisfaction surveys, etc. The latter we can call data informed judgments. With a situative judgment we want to focus on every teacher's continual observation of his students during the teaching. This judgment also takes place on a professional basis but is more holistic and intuitive in that factors such as mood and bodily and verbal expressions, which are not the focus of the teaching, also play a role. It is about quick reflection in action between many different factors, and therefore subjected to misinterpretations. We believe there is a need for systematic feedback from students which can *supplement*, i.e. challenge and adjust, teachers' situative judgments and contribute to teachers' student view. Data-informed judgment may form a new base for considering didactic choices in differentiated teaching.

The teachers involved in the IDDL project have worked with project-oriented teaching through which students were supposed to acquire four project competencies (process management, investigation skills, production skills and problem-solving) that are related to the concepts of 21st century skills. In project-oriented teaching students work more autonomously. Time and space is not as synchronized as in ordinary classroom teaching. When student activities happen in different localities, in different organizational groupings and staggered in time, the teacher may not always pick up on students' body language if and when they have 'clocked out', feel left behind or do not know what to do. In such complex teaching patterns there is a need for nuanced differentiation and thus a need for expanding the teacher's situative judgment. Our assumption is that instant student feedback might help. The use of Elevbaro as a form of digital *remote*-monitoring of the students may supplement teachers' situative judgment. When we use the word 'monitoring' we do not mean to connote social or academic control or supervision according to standards. The concept of monitoring simply means screen-based observation of students. This supplementary remote-monitoring may even contribute to changing (confirming or disconfirming) teachers' views and expectations of students and to teachers' development of their teaching in general.

2. What is Elevbaro?

2.1. Background, design and assumptions

Both differentiated teaching and more complex teaching patterns inspired the development of a digital student feedback tool in the IDDL project. The idea for the first prototype was professor Jeppe Bundsgaard's (Danish School of Education), but in order to facilitate teachers' use of it, it was necessary to develop an actual 'product prototype' (Ulrich & Eppinger, 2008), which is functionally, technically and layout-wise closer to a final solution. In concentrated form, Elevbaro is a *quick and context-dependent* digital tool with which teachers can systematically collect and display *immediate* student feedback on students' *academic and social participation* in a *mobile manner* and through *wordless ratings of fixed statements* with a view to monitoring students, contributing to educational development that is based on perspectives on inclusion and differentiated teaching.

Elevbaro should be used wisely and with a defined purpose, e.g. daily in conjunction with the teaching within a limited course period. The teacher distributes the statements to the students with a click and eventually by date and time stamps (see figure 1).

Figure 1: Student interface²

The theoretical foundation of Elevbaro is a number of assumptions based in the IDDL project's concept of inclusion and differentiated teaching and in the related scaffolding concept and assessment and feedback theory (cf. Carlsen & Graf, under publication; Graf et al., 2016; Graf et al., 2014, Graf, 2012a, 2012b). Specifically, the tool is based on five fixed statements which are jointly presented and on which students are encouraged to take a stand and consider to what extent they agree through a stepless rating inspired by Learning Rating Scale (cf. Nissen, 2012).

Task consciousness (I know what to do) covers the part of a student's intentionality relating to his understanding of the didactic framework. It is a basic prerequisite for a successful acquisition process in formalized teaching.

Needs consciousness (I get the help I need) concerns a student's consciousness of the relation between autonomy and the need for help from 'others', whether a teaching aid, peers or a teacher. It is a broad concept comprising not only academic help as such but also procedural help (e.g. motivation) and social help (e.g. participation).

Challenge consciousness (I feel optimally challenged) is the part of a student's intentionality which is connected to the understanding the concrete subject matter in relation to his own abilities and to his needs consciousness vis-à-vis his commitment. Does the student feel that there is an appropriate problem to solve?

² The interface displays the name of the student grouping (Demo), the subject (Natur/teknologi), the five statements and a Send-button. The statements are translated below in the same order.

Well-being (I feel good in class³) is the personal and social part of a student's academic participation and contains elements of trust, confidence and well-being in the learning situation. Well-being in formalized teaching is understood as having a functional connection to the purpose of the situation.

Commitment (I do my best) identifies a student's attitude to his academic and social participation. Commitment covers more than the element of intentionality of the task consciousness and the challenge consciousness. Commitment is even more than just motivation. It includes an element of responsibility and therefore becomes a counterweight to e.g. a limitless need for help.

All in all the five statements are interrelated and force students – unlike in typical satisfaction surveys – to balance their answers.

The design of the student interface is based on two considerations. First, access through mobile units (e.g. laptop, tablet or mobile phone) makes the tool flexible in complex teaching patterns. While the feedback situation is independent of location, it is essential that students respond at approximately the same time. Secondly, unlike written and complex reflections, the interface favours a bodily element in the response situation in that students have to respond based on their immediate experience and by wordlessly moving a slider. In that way, the tool is a way of capturing something that is close to students' bodily expressions. Our assumption is that the bodily element and the repetition of the easily read statements help to avoid the exclusion of students who do not possess well-developed writing skills nor an elaborate language code.

The core of the teacher interface is the possibility of monitoring through four types of display. It is a simple representation of the students' ratings which, with 100 units of approximately ten centimetres, can be considered stepless. In two types of display the horizontal axis is simply raised to a vertical axis. The two other types of display visualize minimum, maximum, median and quartiles (see figures 2-4). Both types of display can be seen either on a chosen date or over a period of time.

Elevbaro differs from similar Danish feedback tools. Unlike Learning Rating Scale we do not want to 'measure students' learning' (Nissen, 2012: 28). Elevbaro neither can nor will measure general learning or basic well-being according to standards and for the purpose of output control. We 'measure' only in the sense that we convert and display students' self-reported, *in situ* experiences to numerical values on a 100-point scale. Every individual rating is in principle relative to other statements, to other students and from day to day. There is only a minimum of calculation: maximum, minimum, median and quartile. It makes no sense to create averages or other calculations of the five ratings as one factor, of groups of students or of ratings from different days. The ratings, student

³ Class is in a Danish context a set group of students in the same grade that is stable across the teaching of different subjects.

groups, schools and dates are incommensurable. The collected data cannot support statements like the student community is 'x percent committed' or 'x percent feeling well'. Such statements just impose the impression of objectivity and by pass the overall intention of recognizing the individual student. Elevbaro is meant to monitor students' participation in an instant and context-specific way within the relation of mutual trust between teachers and students. We refrain from decontextualized and external use of the data.

In addition, Elevbaro is distinct from *Klassetrivsel.dk* (www.klassetrivsel.dk), which collects survey-based data one or two times a year from the students, generates sociogrammes and suggests new actions to the teacher (e.g. new formations of groups according to CL-principles). In contrast, Elevbaro does not create prescriptive elements, but only displays students' ratings.

We finally assume that the rating of fixed statements over time will qualify the teachers' and students' use of Elevbaro. Unlike tools with open or often changing questions the use over time supports increasing reliability and validity of the Elevbaro-data. When teacher and students continually work with the statements toward a common understanding of them, it will increase the validity of the 'measurements'. Regular and purposeful use of the tool will minimize uncertainty in the rating situation and increase the reliability of the 'measurements'.

3. Methodology and empirical data

In order to study teachers' and students' use and understanding of the new feedback tool within the underexposed field of student feedback we rely on a case study based on an explorative approach (Thisted, 2010: 83) with different data sources and qualitative analysis. More specifically, we are interested in (1) how teachers and students experience the design and interface of Elevbaro, (2) how teachers integrate the tool into their teaching and evaluation practice, and their didactic reflections on their practice, and (3) how teachers and students interpret the statements and tool-generated data. The research can be characterized both as evaluation of a prototype and as heuristic exploration of practice induced by a new tool. The evaluating parts focus on the design and interface of the tool, and examine especially the understanding of the five statements in order to take a first step to validate them in the long run. We want to identify in which way the statements are valid expressions of what the students conceive and to what degree the tool collects the data in a reliable way. By quality of the tool we mean mainly a user-centred definition of quality, that is, whether and how the stakeholders construct meaning. By stressing this open approach we favour the explorative parts of the study that are concerned with questions of where and how new practices and reflections are emerging through the use of the tool. This examination is mainly centred on the question of how teachers interpret the tool-generated data and use them to see and understand their students and their own teaching. Finally we explore whether the use of Elevbaro facilitates new views of the students and affords the development of new, more differentiated teaching.

The core of our empirical data consists of two cases, that is, two teachers in two different schools who both teach a group of sixth-grade students, and the project-oriented classes they have taught within IDDL in the course of the 2014/2015 school year. The data is made up of video recordings and observation notes in the classroom when Elevbaro was introduced by the teachers and used by the students, as well as interviews. The data was collected once at the beginning and a second time closer to the end of the intervention. All in all there were seven individual semistructured interviews and three semistructured group interviews with students plus four semistructured interviews with the two teachers (Kvale, 1997; Tangaard & Brinkmann, 2010). Interviews thematize aspects like time, frequency, context, purpose and consequences of the use of Elevbaro in the classroom. In particular, the students were questioned about their ratings of the five statements and the concrete conditions required for them to place the slider as honestly and in as nuanced a manner as possible. The interviews were carried out as conversations in which the interviewees were asked the same questions directly and several times from different angles and based on different counterfactual possibilities in order to measure their scope of interpretation. Moreover, the teachers were asked to carry out a data analysis in which the thinking-aloud method was combined with questions for elaboration. The interviews have been transcribed and put together with the other empirical elements in Nvivo and coded according to the following overall open categories from both a student's and a teacher's perspective: purpose and use in teaching, didactic integration, data analysis, including the five statements, the four types of display, consequential actions and sense-making. According to our explorative approach we renounced a more formalistic coding and used the categories primarily as structuring keywords so as to identify variation, dominant meaning and unexpected views (Brinkmann & Tanggaard, 2010: 47). The analytical process combined three readings of the material: the data and categories within each case respecting the time and place of collecting, the data contrasting the teacher's and students' perspective and the data across time and place of collecting through the lens of the categories so as to identify the variation of answers. All in all we have chosen to present our interpretations as condensations of meaning which simultaneously reflect the conceptual basis for Elevbaro and its component parts for analysis control (Kvale, 1997: 192, 202, 205). Finally we need to inform, that it was very difficult to translate the quotes word by word in the interviews from a very spoken and context specific Danish language. We are aware of the possibility that our translation of meaning could lead to misunderstanding. For the same reason we tried to keep quotes at a minimum level.

We have handled our dual roles as developers of the tool on the one hand and researchers of it on the other by making explicit a number of prior assumptions (see section 2.1.), which we use as guidelines in our analysis, and by letting the informants' voices be perceptible throughout the analysis – both as regards the informants' creative interpretations and critical comments and by triangulating different sources of data.

4. How students use and understand Elevbaro

In the following we present our analysis of the students' experiences and interpretations of their use of Elevbaro. We focus particularly on the validation of the five statements.

4.1. Students' rating practice in general

First of all, our observations and interviews show that it *is possible* to integrate the tool in teaching without it 'stealing' time from or disturbing the teaching itself or otherwise diverting students' attention. This is confirmed by students and teachers alike. The individual student typically completes a rating within a minute. Secondly, students seem to take the five statements seriously. A negative assumption that students might respond thoughtlessly or mechanically was not confirmed. Nor did students seem to develop a negative attitude to Elevbaro over time. After a year of using Elevbaro, the students questioned still deny being tired of their teacher asking them to rate the statements. Thirdly, our analyses show that students' ratings are generally placed above the middle axis. That means that students often respond in a positive way (see e.g. figure 2). Many students rate one or more statements entirely or almost entirely with a happy smiley. There may be different explanations behind this practice. It could be inferred, for instance, that the students are generally satisfied with the teaching, that the interaction design has a bias, or that students feel a need to please the teacher. The latter we can discount. However, the students' tendency to mostly respond positively raises the question of whether the midpoint of the axis can serve as an adequate interpretation. In this regard the median will do a better job. Fourthly, and despite the above, the students respond to the statements with *some* variation. Different strategies could be observed. Some students carefully push the slider away from the middle for each statement, trying to position it. Others impulsively push one or more sliders either all the way to the top or to the bottom and afterwards readjust them. We also observed students pushing the slider all the way up and then all the way down and then after some hesitation finding the right position. Some students push the slider back and forth within a small area and finally settle on the most adequate position. This all indicates, as we assumed, that the rating situation has a certain bodily spontaneity to it, while at the same time enabling the students to consider the most appropriate feedback. Overall, we may conclude that the interactional design of the tool creates a rating situation for students that is in accordance with its purpose.

4.2. Students' interpretations of the five statements

Our classroom observations and student interviews support that four of the five statements seem to work according to our intentions, while one statement is interpreted too generally.

I know what to do (Task consciousness)

This statement is intuitively understood by students although it does contain a problem regarding time. Even though it is written in the present tense, students understand it to also apply retrospectively as if they were about to begin a task now. Some students added that prospectively they also know what to do tomorrow, and one student also meant for it to include the homework he has to do at home. The concrete rating depends on the context. One student felt that it is a question of whether the teacher has formulated the task properly, whereas another felt it is a question of whether he had been listening. This confirms that the statement is situated between the teacher's didactic framing and the student's interpretation of it. In general, the statement seems to be a good indicator of the students' degree of intentionality in the situation or on the day. As one of the teachers points out, in task consciousness there is also an element of motivation, i.e. the student's interest in the subject matter. Although the students' ratings are typically on the positive side of the axis, there is a useable variation in their responses.

I get the help I need (Needs consciousness)

Even though students give the impression that they have understood the meaning of the statement, it seems as if they do not take it entirely seriously. There are students who understand the statement only as a case of the teacher answering urgent questions from students. The impression is that they usually ask the teacher for help as soon as a problem arises in a project-oriented context, preventing them from finding themselves in situations where they need help that is worth registering. Also, students seem to have difficulty interpreting a need for help more generally as the teacher's scaffolding of their work. Likewise, they have difficulty seeing help as something that can come from their peers or a teaching aid. Yet, we have seen examples of students indicating a lack of help, successfully influencing the speed of teaching or asking for clarification. On an individual level, too, students have successfully attracted the teacher's attention with a low rating. All in all, the statement works as intended, but there will be a need to work with the statement in order to calibrate possible interpretations. In that light, it will be interesting to look into the relation between the students' ratings and the teacher's or expert's examination of the situation.

I feel optimally challenged (Challenge consciousness)

Students seem to understand the statement reasonably uniformly. They say, for instance, that it has 'something to do with the task', 'whether it is too easy or too difficult', 'it is not too difficult, but I still have to think about it before writing anything down', 'if it is easy, I push the slider all the way down.' While 'all the way down' for some students means down by the sad smiley, others simply push the slider to the middle. Although it is in the context of project-oriented work, in which students in some ways take part

in defining their own 'task', the students seem to have a clear idea that the framing of the task is done by the teacher. Without being asked, one group of students mentions how the teacher is surprised or not surprised, praises or praises moderately or sets expectations according to *the student*, e.g. by demanding a more comprehensive product. The students thus relate the statement both to themselves and to the teacher's differentiating in general. Our interview, however, was not close enough to the teaching for us to be able to test to what extent the statement applied to each individual student.

I do my best (Commitment)

The students answer this question, too, using their intuitive understanding. If they're fooling around, not concentrating or focusing on the task, or simply chatting instead, they have not done their best. Although the students can see that perhaps not all of them are willing to activate the necessary self-awareness, there is no risk of being scolded by the teacher. The students seem to hesitate to automatically push the slider all the way up to the happy smiley and often place it in the middle of the right side of the axis. It depends on whether it is easy or difficult, as one student says, thereby placing the statement in relation to the challenge statement. When he is asked a follow-up question about what he does when it is difficult, he is not in doubt that he will do his best. We also ask whether the student is in a dilemma about not always doing well enough and being able to make more of an effort. One student replies that 'I do my best, but I don't do that 110 per cent of the time'. We assume that the students interpret the statement in accordance with the context.

I feel good in class (Well-being)

This statement is not difficult to answer for the students either. Some students expressed that they assess on the background of their feelings and mood that day, whether they have argued with their parents or whether they are simply having a good or bad day. These students exhibit some variation in placing the slider. This meets our expectations that the assessment is meant relate to the specific teaching context. But it can be observed, however, that most students always push the slider all the way up to the happy smiley. They explain, on the one hand, that they feel well in class because they have good friends and are in a good group (Danish: class). The students' responses indicate a more *general* sense of feeling good in class. In other words, these students do not relate to the concrete teaching, but to long-term issues as having friends or being happy to go to school. There is another critical matter of the statement. The term 'class' refers to the Danish way of structuring the students in learning groups with an average of 21.7 participants (2015). When Elevbaro is used in project-oriented teaching, where there is more group work than classroom-based teaching, the term 'class' in the statement is confusing. It is obvious that this Elevbaro-statement does not work according to our assumptions of context sensitive rating and variation in teaching methods, and must therefore be revised.

4.3. Honesty and student ratings

The students asked in the study do not seem to have any problems with honesty in the scores. The students report that naturally the teacher has told them in advance to answer the questions truthfully. In the interviews the students' immediate explanations were challenged when we first asked them to give reasons for their answers and next asked them contrasting questions, e.g. of the 'what if' kind. A few students reject the idea outright of not answering truthfully. Most of them have not thought about this problem, while a few others admit after some discussion that they are able to answer in such a way that it does not provoke the teacher's attention. The question about well-being in class has been particularly controversial with respect to honesty. Some students reply that there is probably *somebody* – understood to mean the others – who might answer untruthfully. The shift to third person shows that the students do not want to appear dishonest. In a group interview, however, they agreed that there could be situations (e.g. bullying) where they would not tell the truth because their situation could get worse. It is clear that the students' way of handling honesty is influenced by the spirit of trust that the teacher has created with the students. For example: 'We're not told off if we indicate that we haven't done our best'. The students say that it benefits neither the teacher nor themselves if they do not respond truthfully 'because then the teachers will get a wrong impression', 'then they think that...' or 'then he might give us tasks that are too difficult'. It seems crucial that students understand the purpose of and the underlying 'contract' for Elevbaro. The students' attitude to Elevbaro fundamentally changed when they discovered that the teacher can actually see their ratings, and that it is he and not the researcher who needs the data. One student said to the teacher, 'If we'd known it was for you, I would have made more of an effort, then I would have..., oh?!'

Finally, we can conclude that the four statements are intuitively comprehensible to students and seem to capture what we intended, but the question of well-being should be clarified. The students' responses are characterized by an adequate degree of honesty. What is crucial to the reliability of the ratings, moreover, is the foundation of trust in the classroom and that students understand the idea behind Elevbaro and how it works. In order to analyse more precisely the value of Elevbaro relative to the individual learning process, it is necessary to follow more students over a longer period of time.

5. How teachers work with Elevbaro

5.1. Using Elevbaro in teaching

Teachers' use of Elevbaro in their teaching is largely unproblematic and works according to the intentions behind the tool. Apart from a few cases at the beginning, when there were still design errors or log-in issues, Elevbaro is used quickly and effectively in teaching contexts. It is mostly used on a daily basis at the end of class over a number

of weeks. Teacher B explains that Elevbaro makes sense to her and that her students like responding to the statements.

We do not have enough empirical data to say whether Elevbaro is especially useful in project-oriented teaching. Teacher A does not wish to use Elevbaro in his ordinary teaching, explaining that, 'In my view, I think that it [Elevbaro] is [best] when you experiment with new teaching, is everybody with me, right? That's probably the context I'd use them in [the statements]'. Teacher B has tried Elevbaro in both project-oriented and regular teaching and explains that there is little difference.

She also points out that it is useful to be able to assign a barometer with a precise date and time already in the planning phase. Not only does it facilitate the teacher's job in the teaching situation, it also prevents students from responding to the statements prematurely. We have observed that confusion may arise if students have several barometers lying around beforehand, as some students will answer the barometer immediately. In other words, it is important that students answer the barometer at approximately the same time.

5.2. Introducing and working with Elevbaro

We can see a clear difference between students who have received minimal introduction to Elevbaro and students who have been allowed to work more thoroughly and continuously with the rating situation and with the statements. By minimal introduction we mean a short oral explanation of Elevbaro prior to using it. Teacher A says: 'Well, all I've done is, I've told them about it and explained that it's a tool for me, so that I can become better at teaching them' and 'I need you (the students) to think about these questions and reply to them truthfully so that I have some data that tells me where to improve, right?'

In a course with two groups of students in the same grade (Danish: parallel classes), we discovered that the two groups showed different rating patterns. In one course the students' ratings were more varied. They used the whole spectre of the scale. The students in the other course answered in a more uniform way, and the ratings were mostly in the top. The teacher in the first course had an interesting way of working with the students' ratings. She invented an exercise in which the students had to go to the blackboard, where she had drawn a long smiley axis, and physically place themselves or write their names on the line at the point where they felt they belonged in connection with a certain statement. Teacher B explains:

Yes, because physically having to go up there and make a decision about where to place yourself, that has also been a good exercise in..., they have actually ... um ... there is a relatively large spread between their positions on the line, I wonder what the B's [the other 6th grade group, ed.] do in class. That will be fun to find out.

We assume that this exercise creates a group-based awareness, contributing to a calibration of the students' understanding of the statements. This can be explained by the fact that students interpret their ratings anew in relation to how the other students have placed themselves. This extended introduction to Elevbaro seems to reduce the rating's randomness and relativity. Furthermore, we suspect that the exercise counteracts both Elevbaro's built-in bias against individualization. Another advantage of the exercise is that teachers are better able to interpret the ratings.

Teacher B has repeated this exercise several times and also spends a bit of time discussing the group's rating practice with the students. She says:

Yes, because we've worked with this in a differentiated way, how ... um ... how far did everybody come, what has everybody's efforts been like, and then talked about, well, what is to be expected, and what's the reason for it, and then something about classroom management, how could I help them to take part in the library sessions, well, where we did the exercise, and they really liked doing it.

There is definitely a need for us to observe this type of group-based discussion about students' rating practices and the effect on their actual rating practices. We assume that the rating practices not only affect the use of Elevbaro, but also contribute to the students' self-monitoring and development of metacognitive skills in general. Could the use of Elevbaro help the students to develop greater autonomy in relation to their work and learning processes? Nevertheless, this will not work on its own. How to frame group discussions about the Elevbaro and the students' rating practices is a didactic question.

5.3. Elevbaro in the context of other forms of evaluation

Both teachers have of their own accord tried to combine Elevbaro with other forms of evaluation. In teacher A's classroom we observed a situation where students were first asked to evaluate the day's project-oriented teaching (an exhibition preview) based on the criteria 'what was good' and 'what didn't work in the project'. After this written, qualitative evaluation on Padlet the students were to reply to Elevbaro, which is a quantitative form of evaluation. Although it can be argued that the students were somehow prepared to reply to the Elevbaro statements, the two forms of evaluation were not explicitly combined. It is merely a temporal order of activities.

In teacher B's classroom we observed a combination of Elevbaro with qualitative, written reflections, also on Padlet, on the project groups' work processes – how far they have got, how the work has come along, how they have worked together, if they are on schedule and whether they have ideas and plans for the following day. Teacher B:

You decide in your groups if it makes sense to begin with Elevbaro first in order to have those questions, they might be useful to have in order to find out how things

went today and how did I work today. Then you can use the words [statements] in Elevbaro for your discussion afterwards to talk about what did we experience today. Right.

Although the teacher leaves the order entirely up to the students, she effectively argues that they begin with Elevbaro because the rating of the statements is useful to them before their upcoming group work. The teacher explicitly combines purpose (monitoring), object (work process) and criteria (comparable questions) for the two forms of evaluation.

Both teachers mention more than once that they need to be able add qualitative statements from the students to their Elevbaro replies. Teacher B explains that, 'sometimes I need clarifications from [a student] about things, about the reason for what I have said'. In addition to the two teachers, others, too, have expressed a desire to supplement Elevbaro with the possibility of obtaining students' qualitative statements, e.g. via a comments box. We have not yet responded to this wish, as it is uncertain whether it will have side effects that might ultimately interfere with the purpose of the tool.

Although these observations are few, there are possibilities for developing a more systematic approach to the combination of Elevbaro with other forms of evaluation. To that end, the following distinctions will be relevant: Is there a focus on student feedback and/or teacher feedback, academic and/or social development, qualitative and/or quantitative approaches, in an oral, written and/or iconic form? All these different ways of evaluating can be combined with Elevbaro.

5.4. The teachers' use and interpretation of the data

In the following we will go through the different display formats in order to thematize the teachers' interpretation of the data from the students' feedback.

Display of the most recent reply

The first thing that Elevbaro shows is an overview of recent replies to each statement for the whole group (figure 2). Five columns indicate each student's reply, with each statement having its own colour. The horizontal axis shows the students and the vertical axis divides the 100 units into +50 (the happy smiley) and -50 (the sad smiley). The display gives a first important overview of the distribution of the students' responses. By using *mouseover* the teachers can get information about a student's name and about the specific numerical values of the respective responses. It is possible to choose one or more students, one or more statements or another date for this type of display, or to combine these options.

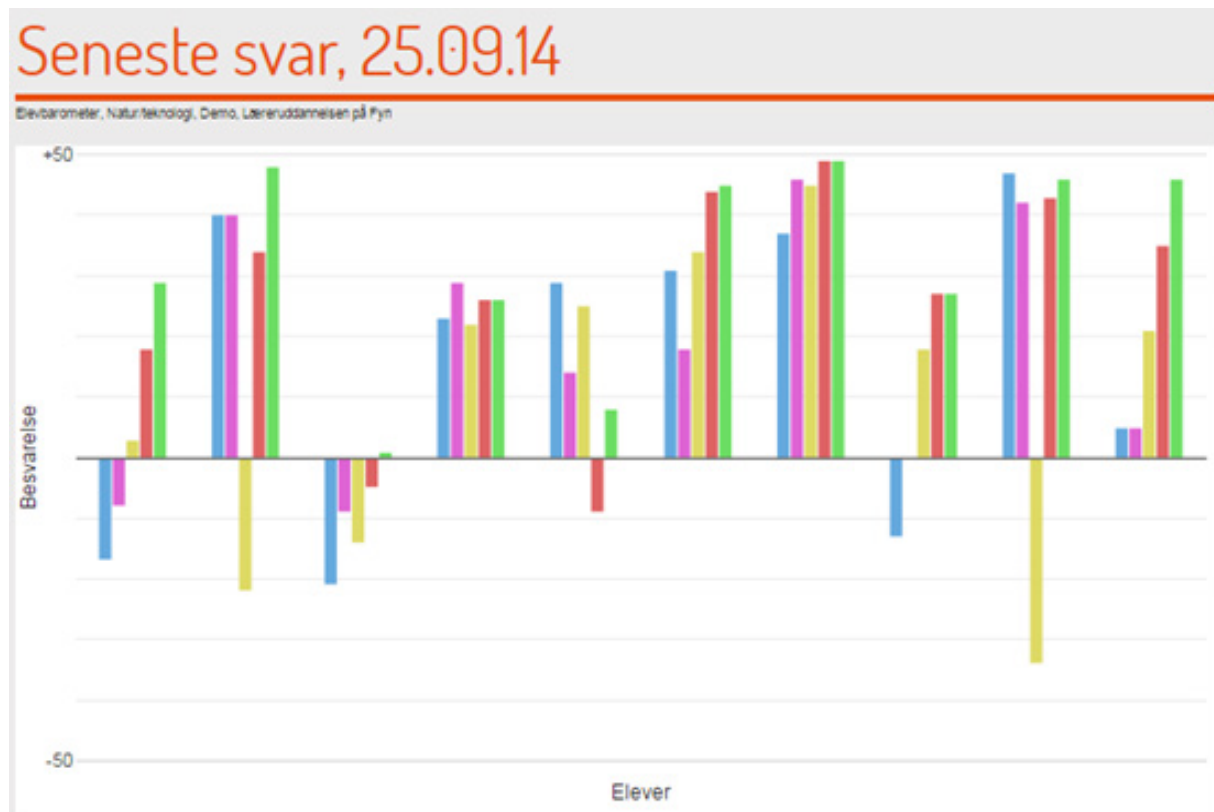


Figure 2: Recent replies with date (Seneste svar, 25.09.14)

The teachers will typically focus first on the replies that fall below the horizontal middle axis. Teacher B says:

Then I typically begin looking at the negative replies here to see what, and then look at, how was their day today, where they are at, so when I see student X here, who hasn't got the help he needed, then I am made aware of that, or make a note of it.

Looking into one group's ratings, teacher B explains: 'When I know them, it makes a whole lot of sense', e.g. student Z, who really wants to take charge of the process and decide things. She explains that student Z is frustrated when he is in a group with two girls he cannot just boss around, but student Y thinks it is fantastic when someone wants to show the way. The teacher's explanation seems to be based on what she already knows about the students.

Display over time

The display over a particular period of time (figure 3) is called 'development' although there is strictly speaking no development but a number of response days which we for practical reasons have connected with a line.

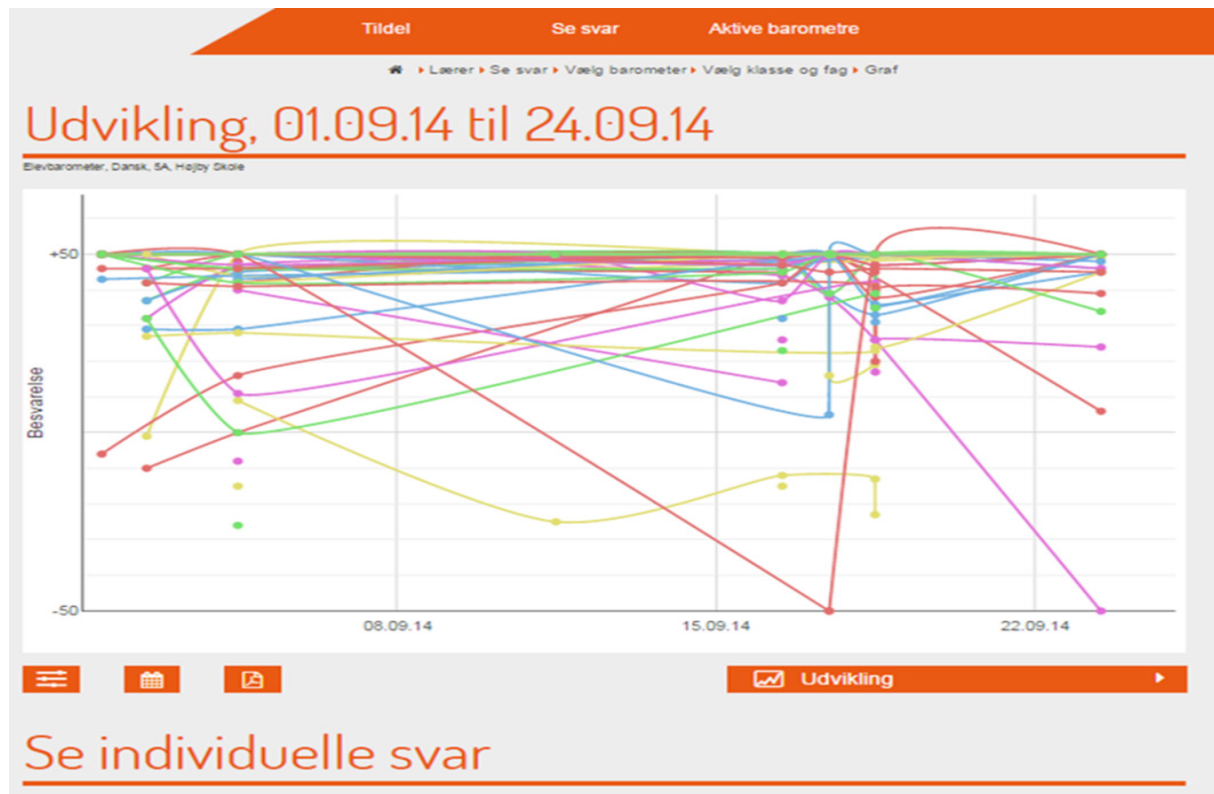


Figure 3: Display over time

This display also seems intuitively useful to teacher B. It makes sense, she says, that there is an upward-moving tendency for the statement about optimal challenge for most students because the first month was rather chaotic, 'but here [pointing at a date] the students became a clearer idea of things'. Teacher B can read the upward-moving change for the statement regarding Task consciousness, as the first feelings of frustration from the study phase have now been replaced with greater clarity. Again, the data analysis fits with her observations of her students and the teaching. The data confirms what she already knew, and does not really challenge her reflections.

Display via box plot

This box plot display is very popular. The display visualizes the variation in the students' responses by means of the span between the maximum and the minimum, the median and the quartiles. It therefore provides a quick overview. 'I know where they are at', as one of the teachers says. The first (blue) box in figure 4, for instance, shows a midrange variation that is relatively close to the 0-midpoint in Task consciousness. It seems that the students didn't really understand the task and there was quite a difference among the students. Here the teacher is somehow invited to reflect on whether there should have been possibilities of differentiation of the teaching. The purple box in contrast shows that most students are very happy with the help they got, so no differentiation is needed.

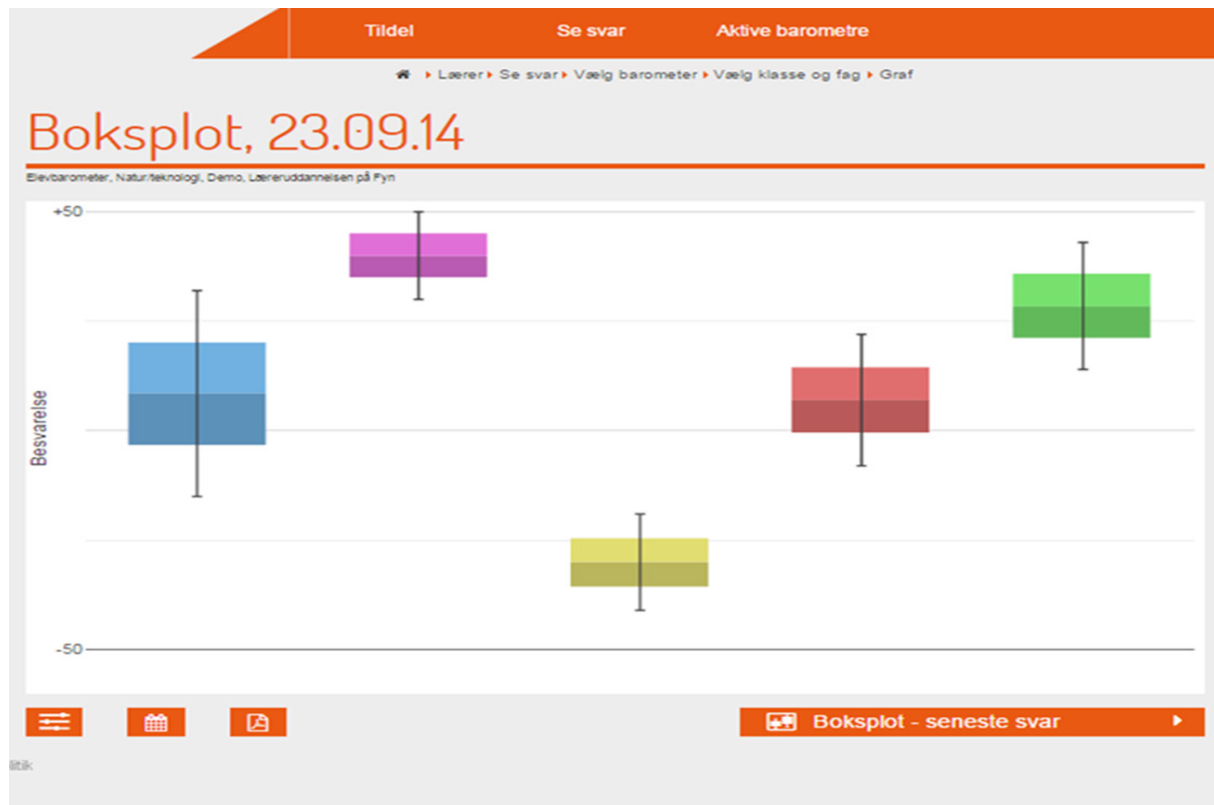


Figure 4: Display via box plot

Finally, there is a fourth display: box plot over time. When using this display, you can follow one of the five statements and see whether the variation in responses changes over time. This display does not seem to be used much, and as we have not asked explicitly about it, we will not discuss it any further here.

Finally we must admit that there will be a challenge for the teachers to use the multiple options for data research. There are undoubtedly limits, however, to how much time teachers can, will and should spend on data analysis.

5.5. Between affirmation and new reflections

Most of the teachers involved seem able to interpret the data they see. Whether the rating from a group of students is generally high, low or mixed, the teachers are able to decode student feedback. For example, teacher A explains that she understands very well why student M replies negatively to the optimal challenge statement because drawing is not his strong suit.

Well, I can see ... that some of them don't give top ratings, which actually fits well with those who do, ... When they give top ratings, it's because they've been active and done a good day's work.

Teacher B also says repeatedly that 'it makes sense' and understands why a certain student responds in a certain way. For example a student who responds negatively to everything because she has been in a bad mood and feels a bit reckless; I know she has done her best; I know that another student did not get the help he needed and that he was not optimally challenged; he is a smart boy, but I didn't manage to get him to set down appropriate challenges for himself during the project work; they're feeling down today because there were problems in their group today; this student probably answers truthfully because she's quite mature; the girl who went home whistling today because she'd worked hard. Some of the assessments can be very complex: the teacher describes the gifted boy, who gives himself a low self-assessment, as a boy who shuns obligations and who has developed avoidance strategies so that we cannot get to him. In this way, the boy often underperforms.

These paraphrased examples indicate, first, that teachers easily recognize students in their responses based on concrete situations and their knowledge of the students. The explanation often stems from a certain type of teaching or activity (frustrating task), from a certain context (their home), from a student's academic skills (not good at drawing) or from the student's mood or personality traits (reckless). The teachers mainly *confirm* their professional image of students. Secondly, the examples show furthermore that the teachers do not question their own teaching in this connection. We could say that the teachers assimilate the data analyses into their existing schemas.

This does not initially support our assumption that the teachers would develop a new outlook. Teacher statements like, 'I wonder why the students were frustrated about my assignment', don't appear often. Nevertheless teacher B acts on the data analysis in a general way. As a consequence of her analysis she declares that she is very observant of all her students, including the quiet ones, and has to keep an eye on them so that they do not disappear completely into a black hole. When we ask her six months later, she believes that the data analysis has enabled her to discover new things. For one thing, she is better at noticing when students are not optimally challenged and follows up by asking students about it. From a teacher presentation about Elevbaro in another context we have a report about a girl who consistently underestimates herself. The teacher does not know how to react, but the data analysis gives her pause for thought. Such examples show where there are opportunities for developing a new outlook. In general we must maintain that the data analysis does not *automatically* challenge teachers' views of their students, although it *is* possible if teachers are focused on it and have the courage to challenge themselves. Surely we need more research into how data analysis may challenge teachers' stable views of their students and their teaching in the long term.

5.6. Teachers' action-taking and sense-making

One thing is quite clear: the data analysis gives rise to renewed attention to one or more students. Teacher B, for example:

When I see student S here, who hasn't got the help he needed, that certainly makes me aware that tomorrow I need to make sure where he's at, so that I can help him...

This type of renewed attention is not strictly a pedagogical or didactic act but rather a step towards it. The digital monitoring leads to renewed face-to-face 'monitoring' that might lead to more concrete actions. In this context, teachers wish to be able to make notes for themselves and send notes to the students through the tool. Teacher B explains that it is difficult to remember everything from data analysis. In this regard, she helps herself, for example, by encouraging a student to remind her of something: 'You need to remind me that we have to talk about what we're working with next time because I can see from Elevbaro that there are things I need to help you with.'

Teacher A finds that it is 'good feedback for me, yeah ... um ... and I might be able to use it in a conversation with the students.' This suggests a pedagogical action in a broader sense. A similar declaration of intent is behind the idea of using Elevbaro for documentation purposes. Teacher A again: 'I want to feel that I'm better at documenting what I'm doing to, well, the head of the school, the parents, the psychologist or whoever I'm talking to.'

Another example is a situation in which teacher B comments on a weak student performance by pointing out that the student has apparently given up during the process and that she has not discovered it. She explains to him that it is *her* responsibility and job to help him, and *his* responsibility to ask for help when he needs it, for instance when he is stuck. She says:

I've seen you sitting at your computer, and it looked as if you were focused on your work ..., and you haven't been very visible, not on Elevbaro either so I could have realised that you haven't had any help in three days, right. Because that's what I use Elevbaro for too, to analyse ... It's my tool. It's precisely so I can keep an eye on those things.

This example shows two things. On one hand the teacher uses the incident to talk about the meaning with Elevbaro. This we characterize as a didactic action in the sense of an action aimed at concrete teaching in this case not in a broader sense, but only in relation to Elevbaro. On the other hand it is obvious that students have to learn when and how to use Elevbaro for it to become a tool for the teacher. In this example we see the teacher enhances the students' metacognition, here about the learning process. We assume that this type of didactic action might have the potential to increase students' continual self-monitoring and thus their autonomy.

We are not able to elaborate on the teachers' actions after data analysis, as we have not observed sufficient subsequent teaching. We hope to study this in another project.

From these interviews and the project consultants' experiences we can finally conclude that teachers can readily see the point in Elevbaro and want to use it more when they

have learned to read the data better. Teacher A: 'I go look at it, and it sort of becomes the mirror I look into when I have to check up on where the kids are at, ... It sort of becomes a mirror that I use for reflection, or a springboard or whatever you call it.'

In conclusion, we may ask whether there is a basis for teachers to rely on their students' feedback on their work and learning processes. From our survey, carried out in the same period with almost 400 teachers in Esbjerg, Denmark, about their current feedback practice, we can conclude that the teachers write off only approximately 10-13 percent of the students when it comes to their self-assessment skills. In other words, teachers have faith in students' self-assessment skills in relation to their academic and social participation, their need for help and their commitment.

6. Conclusions and potentials

We have presented the initial analyses and results from this new way of working systematically with digital student feedback. The testing of the prototype has demonstrated that a minor redesign of the tool is necessary without straying from the original purpose of Elevbaro. We have discussed the basis for the five statements in particular and so far have roughly validated four of them. The question about well-being needs to be revised, however, in order to zoom in on students' more context-specific, social participation. Although the tool is easy to use, didactic integration is necessary on several levels. First of all, a thorough and student-friendly introduction turns out to be essential. If students do not understand the purpose of the tool and its underlying agreement of trust, the students' ratings will not be sufficiently credible and reliable. This is furthermore connected to a collective calibration of individual and relative ratings. Thus and secondly, we identified that the regular rating situation itself and the occasional collective teacher-guided reflections on the rating practice may enhance students' reflection on the teaching and their acquisition process. Therefore we recommend that Elevbaro practices from time to time are thematized in teaching in order to meet the didactic potential for enhanced metacognition and increased student autonomy. To gain more insight into this we need more data both in terms of the number of cases and in terms of a longitudinal approach. Thirdly, it has become clear that quantitative forms of feedback and evaluation like Elevbaro ideally ought to be combined with qualitative forms of feedback and evaluation. Teachers think up ways of combining Elevbaro with other forms of evaluation. Fourthly, teachers can easily make sense of Elevbaro, while time is an issue when it comes to analysing the data. Teachers' reflections in relation to data analysis seem largely to confirm their stable professional student images and their teaching. However, teachers did experience some surprises and were beginning to develop new views on their students. The question of how the Elevbaro can provide an adequate challenge of the teacher's professional judgments still remains. It is obvious that teachers need more competences in data literacy in order to be able to develop their teaching and educational practices by means of Elevbaro.

We think that Elevbaro has promising potential as a tool for learning analytics in the classroom. This must be examined further.

Acknowledgement

We are grateful to Helle Libenholt, lecturer at University College Lillebælt, for her kind assistance with the English language.

References

- Alenkær, Rasmus. (2012). 'Kvalitativ inklusion & IC3', <http://www.alenkaer.dk/resources/kvalitativ-inklusion-og-ic3.pdf>.
- Carlsen, Dorthe; Graf, Stefan Ting. (in publication). Fire differentieringspositioner – teoretisk og empirisk set. In: Bundsgaard, Jeppe; Georgsen, Marianne; Graf, Stefan Ting; Hansen, Thomas Illum; Skott, Charlotte Krog (Eds.). *Innovativ undervisning med it i fagene*. Erfaringer fra tre demonstrationsskoleprojekter. Århus Universitetsforlag, København.
- Carlsen, Dorthe; Graf, Stefan Ting. (2016). *Elevfeedback og Elevbaro – tekniske rapport fra lærersurvey (Esbjerg)*. Læremiddel.dk.
- Dyssegaard, C.; Larsen, M. S.; Tiftikçi, N. (2013). *Effekt og pædagogisk indsats ved inklusion af børn med særlige behov i grundskolen*. Systematisk review. København: Clearing House, forskningsserien, n. 13.
- Egelund, N. (2010). *Undervisningsdifferentiering: status og fremblik*. Frederikshavn: Dafolo. (Vol. 1. udgave).
- Flyvbjerg, Bent. (1991). *Bind 1: Det konkrete videnskab*. København: Akademisk Forlag A/S.
- Graf, Stefan Ting. (2016). *Inklusion og differentiering i digitale læringsmiljøer*: Udviklingsprojekter med demonstrationsskoleforsøg vedr. it i folkeskolen (Slutrapport). Med bidrag fra projektforskere. Odense: Læremiddel.dk.
- Graf, Stefan Ting; Carlsen, Dorthe; Mikkelsen, Stinus Lundum Storm. (2014). 'Inklusion og differentiering i digitale læringsmiljøer: IDDL-projektets fase III og IV: En kompetenceorienteret projektdidaktik og skoleudviklingskoncept i digitale læringsmiljøer'. Retrieved from auuc.demonstrationsskoler.dk/uvd.
- Graf, Stefan Ting. (2012a). Læremidler og almenidaktik – historiske og fænomenologiske ræsonnementer for en ny strukturmodel. In: Graf, Stefan Ting; Hansen, Jens Jørgen; Hansen, Thomas Illum (Eds.). *Læremidler i didaktikken - didaktikken i læremidler*. Århus: Klim i samarbejde med Læremiddel.dk. (p. 61-88).
- Graf, Stefan Ting. (2012b). Læremidler og almenidaktiske modeller – en ny didaktisk strukturmodel. In: Graf, Stefan Ting; Hansen, Jens Jørgen; Hansen, Thomas Illum (Eds.). *Læremidler i didaktikken - didaktikken i læremidler*. Århus: Klim i samarbejde med Læremiddel.dk. (p. 89-114).
- Hattie, John. (2009). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. Routledge.
- Hattie, John; Timperley, Helen. (2013). Styrken ved feedback. In: Andreasen et al. *Feedback og evaluering for læring*. Frederikshavn: Dafolo A/S.

- Hornskov, S.; Berg, H.; Maguire, L. H. (2015). *Review: Brug af data i skoleledelse*. [www.ucviden.dk/portal/da/publications/review-brug-af-data-i-skoleledelse\(2fb3a609-0745-412a-b817-afa892dee50d\).html](http://www.ucviden.dk/portal/da/publications/review-brug-af-data-i-skoleledelse(2fb3a609-0745-412a-b817-afa892dee50d).html).
- Kvale, Steinar. (1997). *Interview*. København: Hans Reitzels Forlag.
- Molbæk, Mette; Tetler, Susan. (2012). 'Klasseledelse med fokus på inklusion og undervisningsdifferentiering'. 4-10-konsortiet. Retrieved from: viauc.dk, juni 2016.
- Nissen, Poul. (2011). Fra erfaringsbaseret til evidensinformeret undervisning. *Kognition og pædagogik*, (79), 21. årg.
- Nissen, Poul. (2012). Hvordan måler man om eleverne lærer noget i skolen? *Økonomi & Politik*, 85(2), 25-33.
- Robinson, V.; Hohepa, M.; Lloyd, C. (2009). *School leadership and student outcomes: Identifying what works and why*. Best Evidence Synthesis Iteration. Wellington: Ministry of Education: New Zealand.
- Brinkmann, Svend; Tanggaard, Lene (Eds.). (2010). *Kvalitative metoder - En grundbog*. København: Hans Reitzels Forlag.
- Thisted, J. (2010). *Forskningsmetode i praksis*. København: Gyldendal Akademisk.
- Ulrich, K. T.; Eppinger, S. D. (2008). *Product Design and Development*. Singapore: McGraw Hill International.
- Yin, R. K. (1994). *Case Study Research: Design and Methods*. London: Sage.

Biographical notes

Stefan Ting Graf has a Ph.D. degree and is senior research fellow at Center for Applied Educational Research at University College Lillebælt, Denmark and working for The National Knowledge Center of Learning Technology. He can be reached on email at stgr@ucl.dk.

Dorthe Carlsen is a lecturer at University College Syddanmark, working for The National Knowledge Center of Learning Technology and a Ph.D. fellow at the Danish School of Education. She can be reached on email at dcars@ucsyd.dk.