

Learning from Assessments and Feedback, A Teacher and Student Perspective.

Piet Koppejan

4708 KN Kalmoesberg 15 Roosendaal
The Netherlands

Abstract

This research is done in an educational organization of a master education. The goal of the research is to find out how students can learn from summative feedback and authentic judging and how assessors can learn how to deliver sustainable feedback and authentic judgments. The research question is 'What is the effect of education of lecturers in making sustainable judgments and delivering sustainable summative feedback on the satisfaction of students to continue their learning development?'

Sub questions:

1. What is the effect of graders in grading summative assessments?
2. What is the effect of education of lecturers in making sustainable judgments so that the satisfaction of students improves?
3. What is the effect of education of lecturers in delivering sustainable feedback so that the satisfaction of students improves?
4. What is the effect of education of students in receiving sustainable summative feedback so that the satisfaction of students improves?

The research design is a combination of a quasi experimental design and quantitative and qualitative research of assessor judgments and delivering summative feedback. In this study It is obvious that judging and delivering sustainable feedback flourish in an environment of professional judging and open dialogues between lecturers, assessors and students. Learning processes of students are of common responsibility of the whole organization and not of a single assessor or lecturer. A system of quality management can support that development to a culture of professional assessing.

- A. Further research in other educational organizations may bring more insight in judging and feedback outcomes. In this study a validated questionnaire for student satisfaction is being designed, based upon criteria of authentic judging and sustainable feedback features. In this research five grading styles of assessors can be determined Low and consistent judging ($m \leq 5,94$) and small scatter ($sd < 0,68$).
- B. Under average and alternate judging ($m \geq 6,0 < 6,5$) and large scatter ($sd > 1,0$).
- C. Under average and consistent judging ($m \geq 6,0 < 6,5$) and moderate scatter ($sd \geq 0,68 \leq 1,0$).
- D. High average and alternate judging ($m \geq 6,5 < 7,0$) and large scatter ($sd > 1,0$).
- E. High average judging ($m \geq 6,5 < 7,0$) and moderate scatter ($sd \geq 0,68 \leq 1,0$).

Introduction

Assessments play an essential role in the education Master Educational Needs (M EN). These assessments lead the student to the professional profile (Andriessen & Manders, 2014) that must be achieved. In learning from assessments there is a focus on agentic engagement (Winstone, Nash, Parker & Rowntree, 2016). In agentic engagement, construction of knowledge (Ponte, 2012) plays a key role. This agentic engagement comes about by cyclical learning (Wakelfield, 2016). Construction of knowledge comes about by means of a cyclical learning process that is based upon research, processing knowledge and learning skills. The learner constructs new knowledge, skills and adjusts attitudes. In the process of agentic engagement the learner gives meaning to those processes. This agentic engagement takes place in a learning environment in which the curriculum offers many choices to the student. The student produces his assessments based upon his choices of the knowledge content and skill development. This is on the one hand a generic content and on the other hand a domain specific content. Dimensions of agentic engagement are: cognitive engagement, emotional engagement and behavioural engagement. The agentic engagement is explicated in the making of the assessments, these assessments are by nature cognitive complex. The key instruments consist of sustainable assessments (Boud & Soler, 2017). Boud and Soler point out that sustainable assessments focus on effective learning for now and in the future. Sustainable learning is a function of what students get out of the education and not how many efforts have been put into the process. In the assessments the student identifies himself as an active learner, identifies

his own level of knowledge and the gaps that are present, practices assessments and judgements; developing skills in time and embodies reflexivity and engagement. The student is in his development the owner of his own learning process. Beck, Skinner and Schwabrow (2011) state that the summative feedback in sustainable assessment also has a formative function in the learning of the student. Assessments contain authentic tasks, do not single restrict to knowledge content, but focus more on learning attitude and metacognition, which include knowledge and social skills, that are needed for a professional. Sustainable feedback, formative as summative must encompass that learning process. In the curriculum are in that order assessments criteria that are derived from the professional profile, formulated holistic and positioned on the highest taxonomic level, cognitive, affective as well as psycho-motor. An example of such a criterion is in Table 1.

Table 1. Example of an Assessment Criterion. Assessment Professional Product 1.

Performance Indicator 2A

By self-direction the student is able to work on the development of socio-emotional competencies and citizenship in complex situations and provide a learning environment that is suitable for the pupils and to the group in appropriate relationship between structure, safety and challenge, taking into account the diversity of stakeholders, accounted from personal professionalism.

The student receives the judgement and summative feedback of the grader of his assessment. The judgement and summative feedback are based upon pre-formulated rubrics, see Table 2. The grader can personalize the indicator by adding his judgement and formative feedback to the indicator.

Table 2. Example Rubric Performance Indicator 2A.

Sufficient till good.

Your communication is sufficiently (good) underlined in the complex situation and tuned sufficiently on the supportive needs of stakeholders and promotes in that way to an improvement of the complex situation. You are showing that by . . . In your elaboration it is sufficiently clear that the tuning on supportive needs of the stakeholders is built upon an underlining from your reflective researching attitude and follow-up research. You explicate that you have insight in the factors that are determined in these complex situations. You can reinforce this by . . .

The summative feedback received by the student is formative for the next assessment. In the above-mentioned example, a bridge is hit between judgement and summative feedback, and is being personalized for the student. Judgement and summative feedback are being supplemented and explicated on a personal student level which feedback, which feed-up and which feed forward is being delivered. It is important for the cyclical learning process of students in this way of learning that the manner of judging and delivering feedback of assessors is alike and fits one another. Students report however great differences between their received judgements, assessment criteria, and feedback on their assessments, the summative feedback. These findings do not only appear to be a thorny issue in this master education, but also are being reported in the research literature (Fyfe, Fyfe, Meyer, Ziman, Sanders, & Hill, 2014; Boud, 2013). This research is about resolving this feedback dilemma.

Learning from summative feedback in a context

In Higher Education (HE) assessment is an important instrument in the curriculum. Assessment steers learning and that is why learning and education eventually are focused on the assessment at the end (Fyfe et al., 2014). All assessments in the M EN are based upon the professional profile to be gained, so there is a strong case of constructive alignment (Biggs, 2003). In constructive alignment the characteristics of the students, the learning goals, the educational activities, the learning results and the assessment criteria are framed on one another and students are being stimulated to develop and use high learning processes. Reflexivity and metacognition play an essential role in that learning process (Boud & Soler, 2017). Enhancing the learning process of the student feedback, summative, following the assessment is considered powerful and formative, (Donovan, 2014) as to shape the process of learning. In curricula in HE in which is being learned from out a socio-constructivistic point of view, the active role of the student is conditional. Reeve and Tseng (2014) introduce the concept of '*agentic engagement*'; the student is product and producer of his own knowledge and skill development. It is reported from different researchers that the students do not at all or partially engage in the received feedback (Rudolph, Foldy, Robinson, Kendall, Taylor, & Simon, 2013). Black and Wiliam (2009) formulate a concrete solution in five key aspects which take place in education to stimulate formative feedback: 1. Learning goals and criteria of success are explicated and shared; 2. Proof of learning is gathered by effective discussion about learning tasks, activities to design; 3. Teachers give feedback that enhances learning; 4. Learners are being activated to be a learning source for each other; 5. Learners are activated to see themselves as owner of their learning process.

The process of delivering judgements and summative feedback on assessments, must connect to the culture of engaging of sustainable feedback. Students can be taught to engage more in the received feedback (Winstone & Nash, 2015). Winstone and Nash have developed a 'feedback-toolkit' for that purpose.

The feedback-toolkit offers students in a very concrete way where feedback is about and how they can process feedback. Feedback is an interactive phenomenon, in which teacher and student both play an active role. On the one hand at first there is the teacher who delivers formative feedback during the learning process (Hattie & Timperley, 2007; Nicol & MacFarlane, 2007) and secondly as assessor adds judgements based upon assessment criteria and delivers summative feedback. On this point Sadler (2009) indicates a lot is to be gained. Assessors seem to have their own reference frame, even if explicit assessment criteria are at hand. In this research assessor behaviour in grading will be analysed.

On the other hand there is the student who receives and processes the feedback in enhancing his development to a higher level in order to obtain the required professional profile, the required final qualifications. By organizing feedback as an active process, the effect of feedback on learning can be sustainable. To that perspective a special place is taken in for feed forward (Hattie & Timperley, 2017; Nicol, 2010). Carless (2013, p.113) defines sustainable feedback as a process of active participation in dialogue. Dunworth and Santiago Sanchez (2016) indicate that in written feedback the dimensions, affective or interpersonal, orientating and transformational, determine the quality of written feedback. These dimensions are helping the learner to generate the feedback out of the social context as in an ongoing process of development to an autonomous self-regulating learner.

Research findings indicate that assessors think of themselves as giving clear judgments and delivering good summative feedback (Mulliner & Tucker, 2016), while students report in research of effects of feedback that they are less positive about judgements and feedback (Mirador, 2013). There is a feedback dilemma. This dilemma can be the cause of interrupting totally or partially the cyclical learning process. On the receiving side the student must be taught to engage more in his feedback (Winstone & Nash, 2015). On the delivering side the teacher must be taught how and in which way he must deliver sustainable summative feedback (Nicol, 2010) in order to set the student in the learning mode. Mulliner and Tucker (2016) conclude in their research on this dilemma that education and intervision, colleagues can help each other to enhance their assessor competencies and delivering sustainable summative feedback. This research will place interventions reacting to the conditions to improve qualitative judgements and summative feedback on the assessor side with the goal to enhance the satisfaction of judgments and summative feedback of students on the other side. And will react upon aspects of feedback to guide students to engage feedback in a better way. For that reason, both perceptions of assessors (Koppejan, 2017) and students will be measured.

II. 2.1 Research question and subquestions

'What is the effect of education and intervision of assessors in sustainable judgements and delivering sustainable summative feedback on the satisfaction of students in continuing their development?'

Sub questions:

1. What is the effect of graders in grading summative assessments?
2. What is the effect of education of lecturers in making sustainable judgements so that the satisfaction of students improves?
3. What is the effect of education of lecturers in delivering sustainable feedback so that the satisfaction of students improves?
4. What is the effect of education of students in receiving sustainable summative feedback so that the satisfaction of students improves?

Hypothesis 1

Assessors differ in assessing based on differences in their reference frame.

Hypothesis 2

Assessors differ in assessing based on differences in the amount of used words in judgement and summative feedback.

Hypothesis 3

Education in assessing is a positive predictor of giving sustainable judgements by teachers.

Hypothesis 4

Education of sustainable summative feedback is a positive predictor of sustainable feedback.

Hypothesis 5

Having dialogues with students in engaging sustainable feedback is a positive predictor of student satisfaction of their development.

Hypothesis 6

Receiving feed forward is a positive predictor of student satisfaction of their learning development.

Method of research

This research is taking place in a master education (HE). In this education fifty lecturers are involved, also in the role of assessor and 1500 master students. In the developing of this education a lot of effort has been put in educating students in self-regulation and making choices in content and reflection on their own professional actions. Also, many efforts are put on enhancing interrater reliability between the lecturers as assessors. For this purpose the features of feedback Nicol (2010); Mulliner and Tucker (2016); Winstone and Nash (2015) are learned by means of dialogue and intervision. This research is done on a moment of ongoing development in that area.

In answering the first sub question a randomized sample ($N = 780$) of judgements and summative feedback has been taken, in which 52 assessors are present. Of each assessor ten to fifteen judgements and summative feedback are analysed.

The research concerning the other sub questions (2, 3, 4) is designed as a quasi-experiment. The whole team is receiving an education in judging and delivering sustainable summative feedback. In the same period sixteen assessors (the experiment group) receive an education, by supporting to their needs of development (O1), education in judging and delivering summative feedback. In the education delivering summative feedback will be taught according to the recent findings of Hattie and Timperley (2007); Nicol(2010); Mullinerand Tucker (2016); Winstoneand Nash (2015); Dunworth& Santiago Sanchez (2016).The other part of the lecturers do not engage in this education. In the experimental group of students, students will be taught according the feedback-toolkit of Winstoneand Nash (2015).From an ethical point of view all lecturers will be educated in the same way of the experimental group. This research will be done in the first semester of the education and in the second semester all lecturers will be educated accordingly.

The design of the research is as follows:

Experimental group:	X1	O1	O2	O3	X2
Control group:	X1	O1			X2

X1 = baseline assessment of the perceived satisfaction of students on the received judgments and summative feedback

O1 = team training of lecturers in judging and delivering of summative feedback

O2 = tuned educating experimental group

O3 = students engaging the feedback-toolkit

X2 = subsequent measurements of the perceived satisfaction of students about receiving judgements and summative feedback by lecturers.

III. 3.2 Instruments

IV. In the education MEN all educational activities are being evaluated by lecturers and students, including the judgements and summative feedback. For this evaluation a short questionnaire is being developed based on the findings of Nicol and MacFarlain-Dick (2006); Hattie en Timperley (2007); Sadler (2009); Nicol (2010)and Mulliner& Tucker (2016). From these findings in dialogue with lectures and students a six-item questionnaire has come about. To determine the construct validity of the scales Short Questionnaire Judgment and Summative Feedback and the Extended Questionnaire Judgment and Summative Feedback a confirmative factor analyses has been conducted to check whether the model fits. In the factor analyses values of Kaisers Mayer Olkin between 0,50 and 0,70 apply as average, between 0,70 and 0,80 as good, between 0,80 en 0,90 as very good and above 0,90 as excellent (Hutcheson & Sofroniou, 1999). After determination of the scales, reliability per scale is determined and is being analysed if a scale is being made more reliable by deleting items.

The short questionnaire contains of a sub scale Judgment (two items) and a sub scale Summative Feedback (three items), see Table3.

Table 3. Short Questionnaire Judgement and Summative Feedback

Items	Score van de Items
How was your assessment graded?	Code 1 – Code 2 - 4 – 5 – 6 – 7 – 8 – 9 – 10
<i>Subscale Judgement</i>	
In the judgement an equal ground is made between the performance-indicators.	Totally equal 5 - 4 - 3 - 2 - 1 Not equal at all
It is clear to me that the judgement has used the learning outcomes of the modules.	Completely clear 5 - 4 - 3 - 2 - 1 Not clear at all
<i>Subscale Summative Feedback</i>	
To which extent do you find the feedback appropriate to the given grade?	Totally appropriate 5 - 4 - 3 - 2 - 1 Not appropriate at all
To which extent does the feedback inform you about processing for the next assessment?	Very good information 5 - 4 - 3 - 2 - 1 Absolutely no information
Give additions, remarks or suggestions about the received feedback on your assessment (no names).	Open field.
For this judgement and summative feedback as a whole I give an appreciation of.	Very satisfactory 5 - 4 - 3 - 2 - 1 Very dissatisfactory

The questionnaire starts by the question of the grade. The height of the grade may be of influence on the satisfaction of the student. The scores Code 1 means that the student extended the maximum words size of the assessment. The score Code 2 (Ephorus-check) means that the student has cited to much out of other work. In the subscale Judgement two questions relate explicitly to the performance indicators and learning outcomes of chosen content by the student. Sadler (2009) points out that assessors not always judge from out the assessment criteria. Nicol (2010) calls this contextual feedback. The subscale Summative Feedback asks after the tone of the feedback (Nicol, 2010; notjudgemental; balanced) and is asked for feedforward (Hattie & Timperley, 2007; Nico, 2010; foreseeing to the next assessment). This online questionnaire is sent to all students after having received the judgement in an assessment period. The Cronbach's Alpha for this questionnaire is ($N = 515$) is 0,91, which is excellent. All items load high on this scale, no items are deleted. This questionnaire is being used as basement measurement and final measurement in this research.

Intervention I (O1) consists of educating the whole staff of lecturers in making sustainable judgements and delivering sustainable summative feedback, a process of calibration.

Intervention II (O2) lies in extension of intervention I. In this intervention II a quasi-experimental design will take place with sixteen assessors in a deepened intervention focused on individual tuning of their assessor and feedback skills. Intervention II consists of supporting assessors in transparent judging according to the assessment criteria at hand and rubrics made from out the theory of valid reliable and transparent judging (Baartman & Kloppenburg, 2013) and delivering sustainable summative feedback (Nicol, 2010) and fostering the emotional intelligence of students according to Winstone and Nash (2016). The basement measurement for intervention II equals the basement measurement of the control group. The follow-up measurements are done after the assessor has received feedback on former judgements. Intervention II contains the features of transparent and careful judging (Sadler, 2009), in which is being learned to hold strictly to the assessment criteria (Vivekananda-Schmidt, MacKillop, Crossley & Wade, 2013), considering that the student may read the judgement differently (Winstone & Nash, 2016) and accounting the characteristics of sustainable summative feedback (Nicol, 2010) and 'agentic engagement' (Reeve & Tseng, 2014). For that reason, judging according to assessment criteria, delivering sustainable feedback and 'agentic engagement', there will be a special focus on feed forward in this research.

4. Data-analyses

Per research question the variables and measure values are pointed out.

'What is the effect of education of lecturers in making sustainable judgements and delivering sustainable summative feedback on the satisfaction of students to continue their learning development?

Subquestions:

5. What is the effect of graders in grading summative assessments?
6. What is the effect of education of lecturers in making sustainable judgements so that the satisfaction of students improves?
7. What is the effect of education of lecturers in delivering sustainable feedback so that the satisfaction of students improves?

8. What is the effect of education of students in receiving sustainable summative feedback so that the satisfaction of students improves?

Sub question 1

Independent variables are: gender on nominal level, on ordinal level: age. The dependent variables are: 'Student', 'Type of assessment', 'Assessor', on a nominal level. On a ratio level, 'Grade' and 'Amount of Used Words'.

Sub question 2

Independent variables are: gender on nominal level, on ordinal level: age. The dependent variables are: 'Grade', 'Judgement by performance indicators', 'Sustainable Summative Feedback' and 'Satisfaction of the students', all on a ratio level.

Sub question 3 and 4.

Independent variables on a nominal level are, gender. The dependent variables are 'Judgement by performance indicators', 'Sustainable Summative Feedback' and 'Satisfaction of the students', all on a ratio level.

In order to answer the research question and the subquestions a multivariate regression analysis is carried out. In this regression analyses causal relation is searched. Therefore the nominal and ordinal variables as independent variables are transformed into dummy variables. By means of a correlation analyses between the explanatory variables the mutual relation will be determined if there is the case of multicollinearity. If the multicollinearity is $\leq 0,50$ than one of the variables will be left out of the model. By ANOVA the most significant model will be chosen.

5. Results of the research

Table 1 shows the results of a randomized sample taken out of a vast dossier of assessment judgements. Per assessor a number of ten to fifteen assessments are taken for the analyses.

Table 1

Results of All Assessments

Assessment	N	Min	Max	M	SD	%
Assessment A	258	4	9,0	6,6	0,94	33,5%
Assessment B	147	4	8	6,55	1,06	19,0%
Assessment C	161	4	8	6,19	1,03	20,8%
Assessment D	207	4	9	6,32	1,04	26,7%
All Assessments	773	4	9	6,44	1,02	100%

Assessors $N = 52$

The results in Table 1 show that the average of all assessments is 6,44. The Kolmogorov-Smirnov test indicates that the results are normally distributed. The ANOVA shows that there are no significant differences between the type of assessments.

Table 2 shows the ANOVA of the variable 'grade' of all assessments.

Table 2

*ANOVA Grade*Assessor*

	Sum of Squares	df	Mean Square	F	p
Between Groups	115,329	51	2,261	2,362	0,000
Within Groups	691,317	722	0,958		
Total	806,646	773			

Assessors $N = 52$; Assessments $N = 773$.

The ANOVA in Table 4 shows that differences in grading are explained by the variable 'Assessor'.

Table 5 shows the results of the basement measurement of the satisfaction of the students of Satisfaction Judgement and Satisfaction Summative Feedback.

Figure 1 shows the used wordsize and used feedback features of the feedback and judgement reports. The average used wordsize is $m = 548,5$. The feedback features are normally distributed. There is no indication that grading style correlates with used feedback features or used wordsize.

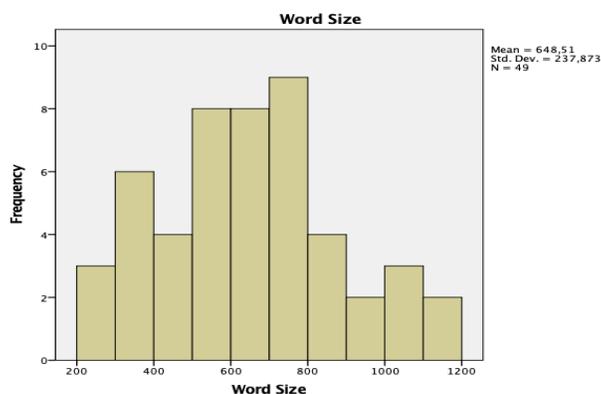


Figure 1 usedWordsize and used feedbackfeatures

Table 5. Results Students Satisfaction, Basement Measurement

Subscales	<i>N</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>
Satisfaction Sustainable Judgment	186	1	5	3,87	1,01
Satisfaction Sustainable Summative Feedback	186	1	5	3,72	1,08
Satisfaction Feedforward	186	1	5	3,72	1,22

The results in Table5 show that students are on average ample satisfactory about the judgement and feedback of their assessment. The Kolmogorov-Smirnov test indicates that the results are not normally distributed. In the student satisfaction the differences are large, $SD \geq 1,0$. Minimum scores of 1 are scored, the student is extremely dissatisfied, and maximum scores (5), the student is extremely satisfied. This is a confirmation of the remarks that students make, there are large differences between the judgements and feedback of assessors. ANOVA in Table6shows the extent of differences between the satisfaction of the students.

Table 6. ANOVA, Differences in Student Satisfaction of Judgements and Summative Feedback

Subscales		<i>Sum of Squares</i>	<i>df</i>	<i>Average Square</i>	<i>F</i>	<i>p</i>
Satisfaction Sustainable Judgements	Between Groups	73,425	6	12,237	19,216	,000
	Within Groups	113,993	179	,637		
	Total	187,418	185			
Satisfaction Sustainable Feedback	Between Groups	61,249	6	10,208	11,740	,000
	Within Groups	155,649	179	,870		
	Total	216,898	185			
Satisfaction Feedforward	Between Groups	41,207	6	6,868	5,247	,000
	Within Groups	231,657	177	1,309		
	Total	272,864	183			

The ANOVA in table 6 shows significant differences on both scales in satisfaction of the judgments and the summative feedback and feedforward.

Table 7 shows the perception of the assessors on assessor competencies and delivering of summative feedback (Koppejan, 2017).

Table 7. Perception of Assessors, Basement Measurement

Subscales	<i>N</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>
Assessor Competencies	33	1	5	4,23	,78
Quality of Summative Feedback	33	1	5	4,00	,69
Quality of Feedforward	33	2	5	3,84	,53

The results in Table 7 show that this group of assessors score themselves on a good level. The results show that there is a scatter in perception ($SD = 0,78/0,69$). The perception on Quality of Feedforward has a small scatter ($SD = 0,53$). These results of the lecturers on competencies and quality of summative feedback and feedforward are corresponding to the results of the student satisfaction on these topics. The difference is not more than half a point.

Table 8 shows the average scores of the experimental group.

Table 8. Results Student Satisfaction of Judgements - Experimental Group Assessors

Subscales	<i>n</i>	<i>Min</i>	<i>Max</i>	<i>m</i>	<i>sd</i>
Satisfaction Judgements	112	1,0	5,0	3,57	,89
Satisfaction Sustainable Feedback	112	1,3	5,0	3,81	,78
Satisfaction Feedforward	112	1,0	5,0	3,61	,80

Note. Assessors $n = 16$

The results of the satisfaction of students given on the assessors of the experimental group are little bit lower than the control group, however there is a smaller scatter ($sd < 0,89$). There are no differences within the experimental group.

Table 9 shows the ANOVA of the control-group and the experimental group.

Tabel 9. ANOVA, Control-group and Experimental Group

<i>Subscales</i>		<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>p</i>
Subscale Judgments	Between Groups	8,905	1	8,905	9,249	0,003
	Within Groups	246,497	256	0,963		
	Total	255,403	257			
Subscale Summative Feedback	Between Groups	0,005	1	0,005	0,944	0,452
	Within Groups	254,053	256	0,992		
	Total	254,058	257			

Table 9 shows that the control-group differs significantly from the experimental group on the subscale 'Judgments'. De control-group has a better score on the subscale 'Judgments' than the experimental group. On the subscale 'Sum

mativ Feedback are no differences between the both groups. Table 10 shows the Effect sizes ($E = \frac{\mu_2 - \mu_1}{\sigma_1}$) of the experimental group. Effectsize greater than 0,2 may be considered as good. Table 10 shows that there is a small effectsize on the scores of All judging items (0,52), on All Feedback items there is a small effectsize (0,18). On the total Evaluation of Feed forward there is an effectsize of 1,00, this is good.

Table 10: Results 0-measurement, measurement 1 and measurement 2

<i>Items</i>	<i>0-measurement</i>		<i>measurement 1</i>		<i>measurement 2</i>		<i>E</i>
	<i>m</i>	<i>sd</i>	<i>m</i>	<i>sd</i>	<i>m</i>	<i>sd</i>	
Total Score All Judging Items	3,77	0,99	3,23	1,07	4,00	0,00	0,52
Total Score All Feedback items	3,72	1,01	3,58	0,64	4,43	0,40	0,18
Total Evaluation Feedforward	3,64	1,15	3,03	0,81	3,78	0,19	1,00

$N = 242$

6. Discussion of the Results

6.1 Reference frame of assessors

The ANOVA Grade*Assessor shows that the results of the grades between assessors differ significantly $F = 2,362$, $p < 0,000$. On further analyses of these data and by comparison of means and standard deviations, by means of an inductive procedure, categories can be made of characteristics of assessor judgements. These characteristics are based upon the means and standard deviation of scores in grading of the assessments of the assessors in this sample. For the category 'means' ($M = 6,44$) the measure is chosen of half a standard deviation ($SD = 1,0$; $1,0 \times 0,5 = 0,5$). This measure results in the following sub categories:

1. Low judging – $m < 5,9$;

2. Under average judging - $m \geq 5,9 < 6,5$,
3. Above average judging - $m \geq 6,5 < 7,0$;
4. High judging - $m \geq 7,0$.

Concerning the measure of scatter ($SD = 1,0$) the choice is to divide this measure into three sub categories. This results in the following sub categories:

5. Small scatter - $sd < 0,68$;
6. Moderate scatter $sd \geq 0,68 \leq 1,0$;
7. Large scatter $sd > 1,0$.

In combining these two sub categories five grading styles of judgments are proposed:

- F. Low and consistent judging ($m \leq 5,94$) and small scatter ($sd < 0,68$).
- G. Under average and alternate judging ($m \geq 6,0 < 6,5$) and large scatter ($sd > 1,0$).
- H. Under average and consistent judging ($m \geq 6,0 < 6,5$) and moderate scatter ($sd \geq 0,68 \leq 1,0$).
- I. High average and alternate judging ($m \geq 6,5 < 7,0$) and large scatter ($sd > 1,0$).
- J. High average judging ($m \geq 6,5 < 7,0$) and moderate scatter ($sd \geq 0,68 \leq 1,0$).

Table 3 shows the numbers distributed among these styles of judgements of the assessors.

Table 3 Distribution of Styles of Judgement of Assessors.					
Style of Judgement	A $m < 5,94$; $sd < 0,68$	B $m \geq 6,0 < 6,5$; $sd > 1,0$	C $m < 6,5$; $sd < 0,68 \leq 1,0$	D $m \geq 6,5 < 7,0$; $sd > 1,0$	E $m \geq 6,5 < 7,0$; $sd < 0,68 \leq 1,0$
N	5	17	8	7	15
	10%	33%	15%	13%	29%

$N = 52$

Table 3 shows that two third of the assessors belong to two styles of judgment, (B) under average judging and large scatter and high (E) average judging and moderate scatter.

In this sample the variable 'Expertise of Assessor' has been included.

Table 4 shows the results of the ANOVA of the variable 'grade' and the mentioned variables.

The ANOVA Grade*Assessor shows that the results of de grades between assessors differ significantly $F = 2,362$, $p < 0,000$. In "Discussion of the results" this will be further analyzed.

Table 3 shows that two third of the assessors belong to two styles of judgment, (B) under average judging and large scatter and high (E) average judging and moderate scatter.

Recommendations and Further Research

The outcomes of this research may be used to investigate other educations and how judging and sustainable feedback are being processed by assessors and students.

The Research Context

This research is done in an educational context where as the organization strictly monitors quality management. Many activities of lecturers and assessors are evaluated on a regular basis. The evaluation of data lead to educating, holding dialogues with lecturers and assessors . Education of assessors lead to fair and open judging and providing sustainable feedback on a high level. That monitoring results in a professional culture of authentic assessments and fair judging and providing sustainable feedback to students. Nevertheless there is no limited result of the quasi experiment. After all, in that invironment students are amply sufficiently satisfied about their learning processes and assessment outcomes. However that education of assessors has not diminished the differences in grading. Therefore it is better to talk about different perspectives of assessors on student performances. And can the student benefit from those different perspectives on his performance. This theme can be central in the dialogues with the team of assessors and group of students.

In this research the experimental group received an assessor focused support by an expert assessor. This support had a good effect on the experimental group. The control group received a team-focused education by supervision and holding dialogues. It is obvious that the experimental group benefitted Of that extra support by the expert assessor.

References

- Andriessen, D., & Manders, P. (2014). Beoordelen is mensenwerk. Bevindingen over de wenselijkheid en mogelijkheid van een gezamenlijk protocol voor het beoordelen van (kern) werkstukken.
- Baartman, L.K.J., & Kloppenburg, R.T.H.M. (2013). KwaliteitsInstrument Toetsprogramma's (KIT) in beroepsgericht onderwijs. Hogeschool Utrecht.
- Beck, R. J., Skinner, W. F., & Schwabrow, L. A. (2011). A study of sustainable assessment theory in higher education tutorials. *Assessment & Evaluation in Higher Education* (November), 37–41.
- Biggs, J. B. (2003). Teaching for quality learning at university. Buckingham: Open University Press/Society for Research into Higher Education. (Second edition).
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21 (1), 5-31.
- Blair, A., Wyburn-Powell, A., & Goodwin, M. (2013). Can dialogue help to improve feedback on examinations? *Studies in Higher Education*, 37-41.
- Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning : the challenge of design. *Assessment & Evaluation in Higher Education*, volume 38.
- Boud, D., & Soler, R. (2017). Assessment & Evaluation in Higher Education Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 41(3), 400–413.
- Donovan, P. (2014). Closing the feedback loop: physics undergraduates' use of feedback comments on laboratory coursework. *Assessment & Evaluation in Higher Education*, 37-41, volume 39.
- Dunworth, K., Santiago Sanchez, H., (2016). Perceptions of quality in staff-student written feedback in higher education; a case study. *Teaching in Higher Education*, March-2517.
- Fyfe, G., Fyfe, S., Meyer, J., Ziman, M., Sanders, K., & Hill, J. (2014). Students reflecting on test performance and feedback: An on-line approach. *Assessment & Evaluation in Higher Education*, 39, 179-194.
- Koppejan, P. (2017). Professional assessing, requires professional training. *Journal of Education & Social Policy*, 2.
- Krathwohl, D. R. (2010). A Revision of Bloom ' s Taxonomy : *Theory Into Practice*, 41(4), 212–218.
- Mirador, J. (2013). 'I'd like to know why': cultural capital and MA in education students' interpretation of feedback commentaries. *Assessment & Evaluation in Higher Education*, 1-14.
- Mulliner, E., & Tucker, M. (2016). Feedback on feedback practice: perceptions of students and academics. *Assessment & Evaluation in Higher Education*, 1-14, volume 42 .
- Nicol, D. (2010). From monologue to dialogue: improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education*, 35(5), 501–517.
- Ponte, P. (2012). Onderwijs en onderzoek van eigen makelij. Onderzoek met en door leraren. Boom Lemma, Den Haag.
- Rudolph, J. W., Foldy, E. G., Robinson, T., Kendall, S., Taylor, S. S., & Simon, R. (2013). Helping Without Harming: the instructor's feedback dilemma in debriefing—a case study.
- Sadler, D. R. (2009). Transforming holistic assessment and grading into a vehicle for complex learning. In Joughin, G. (Ed.) *Assessment, Learning and Judgement in Higher Education* (pp 45-63).
- Vivekananda-Schmidt, P., MacKillop, L., Crossley, J. & Wade, W. (2013). Do assessor comments on a multi-source feedback instrument provide learner-centred feedback? *Medical Education* 2013: 47(11), 1080–1088.
- Wakefield, C. R. (2016). Agentic Engagement , Teacher Support , and Classmate Relatedness — A Reciprocal Path to Student Achievement.
- Winstone, N., E., & Nash, R. A. (2015). The Developing Engagement with Feedback Toolkit (DEFT). *Transforming Teaching Inspiring Learning*. Higher Education Academy.
- Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2016). Supporting Learners ' Agentic Engagement With Feedback : A Systematic Review and a Taxonomy of Recipience Processes. *Educational Psychologist*, 1–21.