Validation of a (peer) feedback perceptions questionnaire

Jan-Willem Strijbos
Leiden University, Leiden, the Netherlands, jwstrijbos@fsw.leidenuniv.nl

Ron J. Pat-El
Leiden University, Leiden, the Netherlands, rpatel@fsw.leidenuniv.nl

Susanne Narciss
Technical University Dresden, Dresden, Germany, susanne.narciss@tu-dresden.de

Abstract

A core educational goal is to produce self-directed lifelong learners who have acquired or are able to acquire the knowledge and skills necessary for their future professional workplace in a networked world. Constructivist and socio-constructivist theories of learning and instruction recommend using collaborative learning scenarios in all kinds of educational contexts in order to attain this goal. Thus, (networked) collaborative learning receives increasing attention in Higher and Further Education. To empower students as self-regulated and collaborative learners peer assessment and peer feedback are increasingly applied in both networked and face-to-face collaborative learning environments. Thus, the conditions under which peer feedback is beneficial for students' learning are worth to be investigated in detail. While the efficiency of feedback content has received much attention in prior feedback research, students' feedback perceptions have been hardly studied. In addition, students' emotional state can mediate the impact of (peer) feedback on their performance. Yet, each kind of feedback has to be perceived and processed in a mindful way, that is, attended to, understood and interpreted, and finally transformed into a corrective action by the student before it can influence subsequent learning. Since up to now the issue of feedback perception has been a neglected area of feedback research, instruments for measuring feedback perceptions are lacking. The purpose of this paper is thus to describe the development and the structural validation of a feedback perception questionnaire which measures students' feedback perceptions in terms of fairness (FA), usefulness (US), acceptance (AC), willingness to improve (WI) and affect (AF). The sample consists of 1535 secondary education students. They received a scenario in which a fictional student received fictional peer feedback on a writing assignment and this feedback was Concise General (CGF) or Elaborated Specific (ESF). The students' were asked to rate their perception of the feedback as if they had received the feedback themselves. The analyses confirm the structural validity of the five scales. FA, US and AC constitute the joined second order component 'Perceived Adequacy of Feedback' (PAF), which in turn positively predicts willingness to improve (WI) and affect (AF). The scale reliabilities were good to excellent.

Keywords

Feedback, peer feedback, perceptions, survey, psychometrics.

Introduction

In instructional contexts the term "feedback" refers to all post-response information which informs the learners on their actual state of learning and/or performance, in order to help them detect if their state corresponds to the learning aims in a given context (Narciss, 2008). Depending on the actual state of learning, feedback can provide a variety of information. In case of no gap between the actual and intended state feedback can, for example, provide information that confirms goal achievement, correctness of a response, or the achieved level of performance. In case of small or large gaps it can provide more or less detailed information, which can be more or less specifically related to learning tasks or processes.

To describe the feedback type variety systematically, several recent reviews and syntheses of research on feedback adopt a multidimensional view of feedback (Hattie & Timperley, 2007; Narciss, 2008; Shute, 2008). Narciss's view stresses three main facets – feedback content, form, and function – that determine the quality of a

ISBN 978-1-86220-225-2

feedback message (Narciss, 2008; Shute, 2008). Combining these facets allows for a large variety of feedback types for instructional settings, which might have differential effects on feedback perceptions and performance.

Feedback content and perceptions

Widely investigated types of feedback are (a) simple feedback types providing outcome-related information, and (b) elaborated feedback types providing additional information besides outcome-related information. Narciss (2008) developed a content-related classification of feedback components which aims at providing a structured overview on simple and elaborated feedback components. Simple feedback components are knowledge of performance, knowledge of result, and knowledge of the correct response. An elaborated feedback component is dependent on the elaborated information provided, which might address: (a) knowledge on task constraints (provides information on task rules, task constraints and task requirements), (b) knowledge about concepts (provides information on conceptual knowledge), (c) knowledge about mistakes (provides information on errors or mistakes), (d) knowledge on how to proceed (know-how) (provides information on procedural knowledge), and (e) knowledge on metacognition (provides information on metacognitive knowledge).

The question of which feedback content is most efficient (i.e., which has the most beneficial effects on performance) has received much attention in prior feedback research. Within most of these feedback studies, the issues of how learners perceive feedback content, and how the perceptions relate to performance have not been addressed explicitly. Yet, several authors have emphasised the 'mindful processing' of feedback as a critical factor for feedback efficiency (Kluger & DeNisi, 1996; Narciss, 2008). Unfortunately, the results of a large body of feedback research are mixed. Only some studies support the commonsense assumption that elaborated and specific feedback affects performance more positively than concise general feedback (see Hattie & Timperley, 2007; Narciss, 2008; Shute, 2008).

Nevertheless, this common sense assumption is often reflected in feedback perceptions, for example, studies on writing instruction reveal that students perceive feedback providing elaborated and specific advice more positively than concise evaluative feedback (Arndt, 1993). If so, students incorporated peer and teacher feedback in their revisions, whereas they ignored any feedback that was either not specific enough and/or did not explain the problems (Tsui & Ng, 2000). Nevertheless, the perception of peer feedback – if measured – is commonly measured in terms of the single dimension "usefulness", after the feedback has been applied and/or at the end of the task (Kwok, 2008). Furthermore, as shown in many studies on feedback seeking, even the most sophisticated feedback is useless if learners do not attend to it or are not willing to invest time and effort in error correction. Finally, there is also accumulating evidence that students' emotional state can mediate the impact of feedback on their performance (Shute, 2008).

Whereas the feedback content in most cognitive feedback studies was designed, by combining simple and elaborated components, the content of a feedback message by a peer depends on the combination of simple and elaborated feedback components *generated by a fellow student*. Consequently, students' perception of the feedback content could play a large role in their response to peer feedback.

Peer feedback content and perceptions

In instructional contexts there are at least five feedback sources, namely the teacher, peer, parents, book or computer-based environment (Hattie & Timperley, 2007), and the task (if the learning goals and criteria for goal attainment are defined in a way that learners can self-assess their learning). Depending on the source's characteristics, feedback content might be perceived as less useful or less credible, and affect task completion or learning differentially. Whereas feedback by a teacher, a book or a parent represents information from an authoritative source, a peer is not always seen as a reliable source by their fellow students. In fact, students often voice that evaluation is a role of the teacher (Brown, Irving, Peterson, & Hirschfeld, 2009). Nevertheless, recent studies indicate that peer feedback can be associated with a larger degree of student autonomy (Yang, Badger, & Yu, 2006) and teacher feedback can result in passive and dependent learners (Lee, 2008). Yet, as students are not experts in a subject area, peer feedback is susceptible to variation. In addition, students often doubt their own and their peers' knowledge within a given subject area, and their own and their peers' skill to evaluate a peer (Van Gennip, Segers, Tillema, in press). Moreover, peer feedback content can vary due to friendships whereby students prefer not to assess their peer too harshly (Cheng & Warren, 1997). When people feel threatened, e.g. when they are judged or the feedback is not in line with their own view of self-performance, fairness becomes important (Leung, Su, & Morris, 2001) making the acceptance of peer feedback (irrespective of its accuracy) less likely. Peer feedback is increasing applied in online learning contexts, for example during

ISBN 978-1-86220-225-2

the revision of essays (Cho & Schunn, 2007). Furthermore, peer assessment and peer feedback can be applied to both group process and product assessment during collaborative learning in networked learning environments (Phielix, Prins, & Kirschner, 2010).

Research objectives

Up to now the issue of feedback perception has been neglected in feedback research, and as a consequence instruments for measuring feedback perceptions are lacking. Feedback which is perceived as fair and useful might be more attended to, as compared to feedback that is perceived as unfair, and useless. In addition, information about the particular sender of feedback could trigger social comparison, and as result influence how a recipient perceives and treats the feedback. In other words, the feedback perception may affect students' willingness to improve and their affect (emotional state), which may subsequently influence performance. This paper investigates the structural construct validation of a multidimensional feedback perceptions questionnaire. More specifically: (a) Can feedback perceptions be captured adequately and robustly with this multidimensional questionnaire?, (b) Is feedback perception predictive of willingness to improve and affect?, and (c) Is this questionnaire invariant for different types of feedback and gender?

Method

Sample

The sample consisted of 1535 pre-university and senior general secondary education students in the Netherlands from 130 schools. There were 817 female and 713 male students and their mean age was 15.75 (*SD* = 1.19). In each school data was collected in four classrooms, covering all disciplines ranging from arts to sciences, in which four students were randomly selected by their teacher to complete the questionnaire. Participation to the study was based on informed consent. The classroom response rate was 98.40% and the participating schools were spread across different regions in the Netherlands to avoid a bias towards urban areas.

Materials

As part of large scale questionnaire study, students were presented with a scenario in which a fictional student received feedback by a fictional peer. The scenario was embedded in the context of the task of 'writing a business letter'. In addition to the peer feedback the students received the evaluation criteria for a business letter (main criteria: components, content, spelling and style) and the fictional "letter assignment". Two feedback scenarios were designed: feedback content was Concise General (CGF) or Elaborated Specific (ESF). CGF contained solely general remarks regarding the quality of the performance, whereas ESF provided the position and error type, as well as information on how to proceed (see Appendix A).

Measures

We used a multidimensional 18-item feedback perceptions questionnaire (Appendix B), measuring feedback perceptions in terms of fairness (FA), usefulness (US), acceptance (AC), willingness to improve (WI) and affect (AF). Questionnaire items were measured on a 10 cm bi-polar scale from 0 (fully disagree) to 10 (fully agree). Four scales consist of three items (FA, US, AC and WI). Affect was measured with six items; three measuring positive affect and three measuring negative affect. Negatively phrased items were recoded. Items were phrased using English, German and Dutch adjectives – through translation and re-translation – to ensure that items addressed the same semantic aspects. The questionnaire was previously used with higher education students in a German (Strijbos, Narciss, & Dünnebier, in press) and Dutch (Strijbos, Narciss, & Segers, 2009) context.

Procedure

Schools were visited by research assistants who distributed the questionnaires in the classrooms. When presented with the scenario the students were asked to consider the peer feedback, as if they had received the feedback themselves, and indicate how they perceived the peer feedback.

Data analysis

The validation was conducted through confirmatory factor analyses using Structural Equation Modelling (SEM) to determine (a) the robustness of the factor structure reported by Strijbos et al. (in press), (b) whether feedback perception predicts willingness to improve and affect, and (c) whether the factor structure is similar for different types of feedback and gender.

ISBN 978-1-86220-225-2

Confirmatory factor analysis
SEM was performed in EQS version 6.1. To interpret a model's fit, the following indicators were used: SRMR and RMSEA below 0.10 is considered adequate fit and below 0.05 an excellent fit, and CFI scores above 0.90 indicate adequate fit and above 0.95 excellent fit (Browne & Cudeck, 1992). Since the χ^2 statistic becomes increasingly unreliable in sample sizes > 250, the significance of the χ^2 statistic will not be used as a criterion for model fit (Byrne, 2006).
Invariance tests
Increasing levels of equality constraints were applied to the tested models to assess structural invariance of the questionnaires between different types of feedback and gender. The tested types of invariance are (a) dimensional (baseline): is the same number of (common) factors present in each group; (b) configural: are (common) factors associated with the same items across groups; (c) metric: do (common) factors have the same meanings across groups; (d) strong factorial: are comparisons of group means meaningful; (e) strict factorial: are comparisons of group means and observed variables defensible (Gregorich, 2006). A critical value of -0.01 ΔCFI will be used to judge structural equivalence (Cheung & Rensvold, 2002).

Results

Confirmatory factor analysis

Structural equation modelling on all 18-items – with a common factor PAF for FA, US and AC items, and a common factor for all AF items – yielded a very weak fit, $\chi^2(127) = 2752.36$, CFI = .799; SRMR = .189; RMSEA = .116. No Heywood cases – negative estimations of error variances for an indicator of a latent variable – were found in any of the fitted models. We then conducted a separate analysis for the common factor PAF and a separate analysis for WI+AF.

Structural equation modelling was used to confirm the second-order factor structure of PAF (with correlated errors for negatively worded items AC2 and AC3). The proposed second-order factor structure fitted adequately, but a high RMSEA indicated poor fit, $\chi^2(22) = 576.07$, CFI = .929, SRMR = .051, RMSEA = .128. Inspection of LM-multipliers suggested that item FA1 was more indicative of the US scale. This change resulted in a excellent model fit, $\chi^2(22) = 350.52$, CFI = .958, SRMR = .039, RMSEA = .098. The proposed correlated first-order structure of WI+AF was tested. The initial correlated factor model fitted poorly, $\chi^2(26) = 1390.903$, CFI = .612, SRMR = .143, RMSEA = .185. Inspection of the LM-multipliers suggested a positive wording effect in the AF scale due to both negatively and positively worded items. Correlating errors for the positively worded AF items yielded an excellent fit, $\chi^2(23) = 278.26$, CFI = .927, SRMR = .070, RMSEA = .085).

Finally, to test the proposed theoretical relationship between PAF, WI and AF, a path analysis was conducted using SEM. Modelling PAF as a predictor of both WI and AF yielded a good fit, $\chi^2(122) = 1636.07$, CFI = .884, SRMR = .074, RMSEA = .090. Figure 1 shows the path estimates, as well as the first and second order factor loadings. Appendix B shows the variables, items and Cronbach's alpha per scale.

ISBN 978-1-86220-225-2

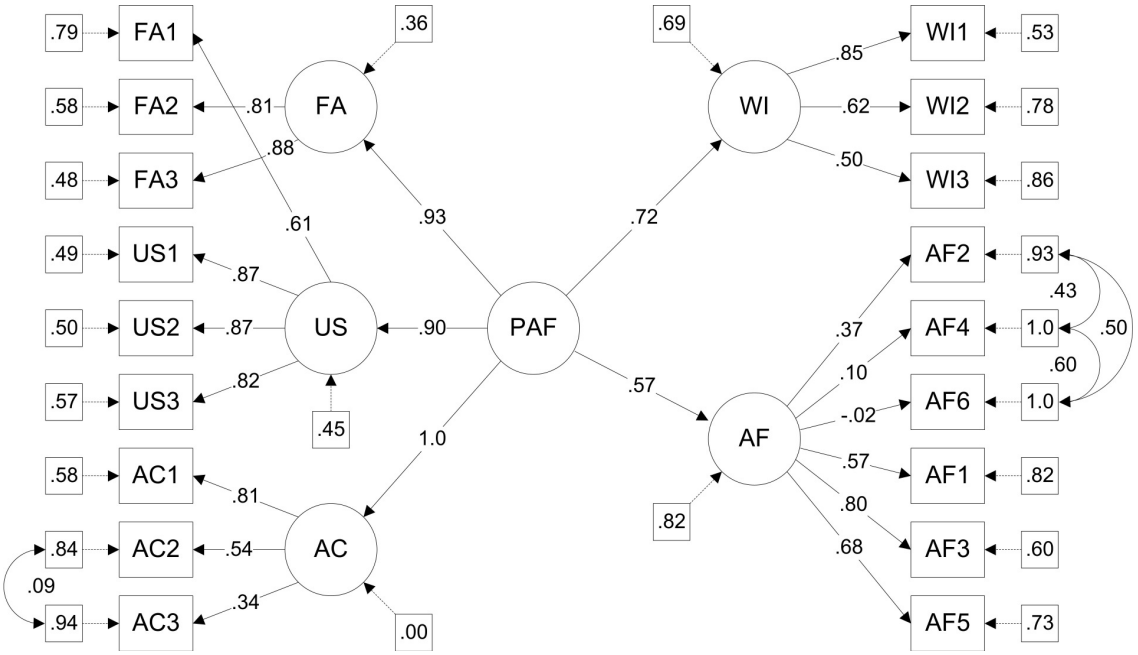

Figure 1. Path estimates and first and second order loadings.

Invariance tests

Tests for measurement invariance were conducted for PAF and WI+AF on (a) type of feedback and (b) gender. Table 1 presents the findings for all tested models.

Invariance for type of feedback

Testing for the baseline hypothesised model for the PAF yielded an excellent fit in the CGF (model F1a) and ESF condition (model F1b). Incremental adding of constraints provided evidence for metric invariance for equal first order (model F3a) and second order loadings and covariances (model F3b), indicating that the factors have the same meanings across type of feedback conditions (in both models ΔCFI = .002). Constraints on observed variable intercepts yielded a well fitting model (model F4a). Testing for the invariance of latent intercepts revealed a well-fitting model (model F4b) providing evidence for strong factorial invariance. Constraints on all estimated error terms did not fit satisfactory (Model F5). Testing for the baseline hypothesised model for WI+AF yielded an excellent fit in the CGF (model F6a) and ESF condition (model F6b). Incremental adding of constraints provided evidence for metric invariance of WI+AF, indicating that the factors have the same meanings across type of feedback (model F8). Constraints on observed variable intercepts yielded a well fitting model (model F9) providing evidence for strong factorial invariance. Constraints on all estimated error terms resulted in a weak fit (Model F10).

Invariance for gender

Testing for the baseline hypothesised model for the PAF yielded an excellent fit for male (model G1a) and female (model G1b) students. Incremental adding of constraints provided evidence for metric invariance for equal first order (model G3a) and second order loadings and covariances (model G3b), indicating that the factors have the same meanings across gender (in both models ΔCFI = .002). Constraints on observed variable intercepts yielded a well fitting model (model G4a). Testing for the invariance of latent intercepts revealed a well-fitting model (model G4b) providing evidence for strong factorial invariance. Constraints on all estimated error terms yielded a good fit (Model G5), providing evidence for strict invariance. Testing for the baseline hypothesised model for WI+AF yielded an excellent fit for male (model G6a) and female (model G6b) students. Incremental adding of constraints provided evidence for metric invariance of WI+AF, indicating that the factors have the same meanings across type of feedback (model G8). Constraints on observed variable intercepts yielded a well fitting model (model G9) providing evidence for strong factorial invariance, but constraints on all estimated error terms did not result in a satisfactory fit (Model G10).

ISBN 978-1-86220-225-2

Discussion

The aim of our study was to (a) investigate the robustness of the factorial structure of a multidimensional feedback perceptions questionnaire, (b) test whether feedback perception was a predictor of willingness to improve (WI) and affect (AF), and (c) investigate to what extent the multidimensional feedback perceptions questionnaire was similarly interpreted across two types of peer feedback – Concise General (CGF) and Elaborated Specific (ESF) – and across gender.

Perceived fairness, usefulness and acceptance appeared constituent parts of 'perceived adequacy of feedback' (PAF). In addition, WI and AF were correlated – yet distinct measures. The factorial structure for the present secondary education sample confirmed the Oblimin principal components reported by Strijbos et al. (in press) and Strijbos et al. (2009). PAF was confirmed as a predictor of WI and AF. Both the PAF and WI+AF part of the questionnaire were invariant across both types of peer feedback as well as gender: the PAF and WI+AF parts are not only robust, but are also similarly interpreted across both peer feedback types and gender. Our finding that corresponding latent factor intercepts and/or observed variable intercepts are invariant across those groups suggests that (a) group differences in estimated factor means are unbiased and (b) group differences in observed means are directly related to group differences in factor means and not contaminated by differential response bias (Gregorich, 2006). For both types of feedback and gender a strong factorial level of invariance was found. For researchers this implies that comparisons of group means are meaningful. Strong factorial invariance is statistically important to meaningfully defend comparisons of factor and observed means. PAF was found to be strictly invariant for gender, which makes not only comparisons of group means defensible, but also of observed variances and covariances. It might be argued that WI+AF could be deemed strictly invariant across types of feedback, because the reduction in fit is only slightly above the determined threshold. The decision to adhere to the $\Delta CFI < .01$ criterion was based on prudence; rather report a robust strong factorial invariance than an unclear instance of strict invariance. Although strict invariance is desirable for comparisons, it is usually considered too stringent (Byrne, 2006), whereas strong factorial invariance is deemed a more readily attainable goal (Gregorich, 2006). Strong factorial invariance is in most cases adequate for comparative research.

ISBN 978-1-86220-225-2

Table 1: Structural equivalence (baseline and configural) and invariance (metric, strong and strict) model comparisons for PAF, and for WI+AF.

PAF								
Feedback (F)	Model	χ^2	df	CFI	SRMR	RMSEA	RMSEA 90% CI	Model comp.	ΔCFI
CGF baseline	F1a	200.45	22	.950	.048	.103	.090, .116		
ESF baseline	F1b	188.05	22	.952	.036	.099	.086, .112		
Configural	F2	388.51	44	.951	.042	.101	.092, .110		
Metric 1st order	F3a	411.12	51	.949	.052	.096	.087, .104	F3a vs. F2	.002
Metric 2nd order	F3b	413.02	54	.949	.053	.093.	.085, .101	F3b vs. F2	.002
Strong factorial 1	F4a	472.00	54	.952	.060	.095	.086, .103	F4a vs. F2	-.001
Strong factorial 2	F4b	722.91	57	.950	.101	.097	.088, .105	F4b vs. F2	.001
Strict factorial	F5	974.80	68	.921	.114	.112	.105, .120	F5 vs. F2	.030
Gender (G)									
Male	G1a	159.15	22	.961	.042	.094	.080, .107		
Female	G1b	222.08	22	.952	.040	.106	.093, .118		
Configural	G2	318.31	44	.961	.042	.094	.084, .103		
Metric 1st order	G3a	389.82	51	.956	.044	.093	.085, .102	G3a vs. G2	.005
Metric 2nd order	G3b	391.30	54	.956	.045	.090	.082, .099	G3b vs. G2	.005
Strong factorial 1	G4a	404.76	54	.956	.045	.090	.082, .099	G4a vs. G2	.005
Strong factorial 2	G4b	413.08	57	.956	.045	.090	.082, .090	G4b vs. G2	.006
Strict factorial	G5	461.63	68	.952	.050	.087	.079, .095	G5 vs. G2	.009
WI+AF									
Feedback (F)	Model	χ^2	df	CFI	SRMR	RMSEA	RMSEA 90% CI	Model comp.	ΔCFI
CGF baseline	F6a	148.00	23	.930	.072	.084	.071, .097		
ESF baseline	F6b	144.39	23	.927	.073	.083	.070, .096		
Configural	F7	292.39	46	.929	.072	.083	.074, .093		
Metric 1st order	F8	328.73	56	.922	.077	.080	.071, .088	F8 vs. F7	.008
Strong factorial	F9	448.53	63	.924	.080	.080	.072, .088	F9 vs. F7	.005
Strict factorial	F10	488.50	72	.916	.084	.079	.071, .086	F10 vs. F7	.013
Gender (G)									
Male	G6a	168.85	23	.906	.071	.094	.081, .108		
Female	G6b	155.09	23	.934	.076	.084	.071, .096		
Configural	G7	319.23	46	.917	.078	.091	.082, .101		
Metric 1st order	G8	326.18	56	.918	.079	.082	.074, .091	G8 vs. G7	-0.01
Strong factorial	G9	371.47	63	.919	.079	.082	.074, .091	G9 vs. G7	-0.02
Strict factorial	G10	492.80	72	.888	.087	.091	.083, .099	G10 vs. G7	.029

The results clearly reveal that students' feedback perception – in terms PAF, WI and AF – can be adequately captured with the multidimensional questionnaire. Given the increased recent interest for formative assessment and feedback practices (between peers or by a teacher) students' perception of feedback could be a crucial determinant of how they treat the feedback and possibly help to uncover when and how feedback perception is positively related to feedback efficiency. This questionnaire enables the comparison of feedback perceptions in networked and face-to-face contexts to explore how the specific nature of these contexts might influence the dynamics and efficiency of peer feedback. Moreover, although anonymous peer feedback might be appealing in online contexts, formative feedback aims to foster reflection and change in performance and collaborative skills; including skills to cope with feedback by non-anonymous peers or colleagues in students' future professional life (Strijbos, Ochoa, Sluijsmans, Segers, & Tillema, 2009).

References

Arndt, V. (1993). Response to writing: Using feedback to inform the writing process. In M. N. Brock & L. Walters (Eds.), *Teaching composition around the pacific rim: Politics and pedagogy* (pp. 90-116). Clevedon, UK: Multilingual Matters.

Brown, G. T. L., Irving, S. E., Peterson, E. R., & Hirschfeld, G. H. F. (2009). Use of interactive–informal assessment practices: New Zealand secondary students' conceptions of assessment. *Learning and Instruction*, *19*, 97-111.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research*, *21*, 230–258.

Byrne, B. M. (2006). *Structural equation modelling with EQS*. Mahwah, NJ: Erlbaum.

Cheng, W., & Warren, M. (1997). Having second thoughts: Student perceptions before and after a peer assessment exercise. *Studies in Higher Education*, *22*, 233-239.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233-255.

Cho, K., & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers and Education*, *48*, 409-426.

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory analysis framework. *Medical Care*, *44*, S78-S94.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*, 81-112.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*, 254-284.

Kwok, L. (2008). Students' perception of peer evaluation and teachers' role in seminar discussions. *Electronic Journal of Foreign Language Teaching*, *5*, 89-97.

Lee, I. (2008). Student reactions to teacher feedback in two Hong Kong secondary classrooms. *Journal of Second Language Writing*, *17*, 144-164.

Leung, K., Su, S., & Morris, M. W. (2001). When is criticism not constructive? The roles of fairness perceptions and dispositional attributions in employee acceptance of critical supervisory feedback. *Human Relations*, *54*, 1155-1187.

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. Van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 125-143). Mahwah, NJ: Erlbaum.

Phielix, C., Prins, F., & Kirschner, P. A. (2010). Awareness of group performance in a CSCL environment: Effects of peer feedback and reflection. *Computers in Human Behavior*, *26*, 151-161.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*, 153-189.

Strijbos, J. W., Narciss, S., & Dünnebier, K. (in press). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction*.

Strijbos, J. W., Narciss, S., & Segers, M. S. R. (2009, August). *Peer feedback in undergraduate academic writing: How do feedback content, writing ability-level and gender of the sender affect feedback perception and performance?* Paper presented at the 13th biennial EARLI conference, Amsterdam, the Netherlands.

Strijbos, J. W., Ochoa, T. A., Sluijsmans, D. M. A., Segers, M. S. R., & Tillema, H. H. (2009). Fostering interactivity through formative peer assessment in web-based collaborative learning environments. In C. Mourlas, N. Tsianos, & P. Germanakos (Eds.), *Cognitive and emotional processes in web-based education: Integrating human factors and personalization* (pp. 375-395). Hershey, PA: IGI Global.

Tsui, A. B. M., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Learning*, *9*, 147-170.

ISBN 978-1-86220-225-2

Van Gennip, N. A. E., Segers, M. S. R., Tillema, H. H. (in press). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction*.
Yang, M., Badger, R., & Yu, Z. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing*, *15*, 179-200.

Appendix A

Concise General Feedback (CGF)		Feedback component
Components	There are errors in technical components.	KR + KM (general)
Content	The content is sometimes too extensive.	KR + KM (general)
Spelling	There are spelling errors in the letter.	KR + KM (general)
Style	Sometimes there are style errors in the letter.	KR + KM (general)

Elaborated Specific Feedback (ESF)		Feedback component
Components	There are errors in technical components, like in the letter head.	KR + KM
	Also the address and date are not written correctly.	KR + KM
	The paragraphs are not neatly aligned, although this should be the case.	KR + KM + KH (implicit)
Content	Everything that should be in the letter is included, it is a bit extensive though.	KR + KM (general)
	Sometimes things are included that do not apply, such as the bank account number.	KR + KM + KH (implicit)
	Marieke does ask not for a financial refund, thus it is not necessary to write the bank account number.	KR + KM + KH
Spelling	There are spelling errors in the letter.	KR + KM (general)
	In the last sentence of the first paragraph for example, the word "voltooid" is written with a **t** and not with a **d**, even though the present perfect applies.	KR + KM + KH (implicit)
Style	Sometimes there are style errors in the letter.	KR + KM (general)
	In business letters, for example, you cannot use the "&" symbol.	KR + KM + KH (implicit)
	You can also not start a sentence with "And".	KR + KM + KH (implicit)
	It is better not to write "Again a disappointing telephone conversation".	KR + KM + KH

Note on feedback components (see Narciss, 2008): KR (knowledge of result/ response), KM (knowledge of mistakes) and KH (knowledge on how to proceed). An additive strategy was used to design the ESF, i.e. CSF + further components listed).

Appendix B

Scale	α	Variable	Item
Fairness	.71	FA1	I would be satisfied with this feedback.
		FA2	I would consider this feedback fair.
		FA3	I would consider this feedback justified.
Usefulness	.89	US1	I would consider this feedback useful.
		US2	I would consider this feedback helpful.
		US3	This feedback would provide me a lot of support.
Acceptance	.55	AC1	I would accept this feedback.
		AC2	I would dispute this feedback.
		AC3	I would reject this feedback.
Perceived Adequacy of Feedback	.89	PAF	FA + AC + US items.
Willingness to improve	.71	WI1	I would be willing to improve my performance.
		WI2	I would be willing to invest a lot of effort in my revision.
		WI3	I would be willing to work on further business letter assignments.
Affect	.67		I would feel … if I received this feedback on my revision.
Positive	.73	AF2	Satisfied
		AF4	Confident
		AF6	Successful
Negative	.73	AF1	Offended
		AF3	Angry
		AF5	Frustrated

ISBN 978-1-86220-225-2