* Title Page WITH author details

Self-confidence and performance goal orientation interactively predict performance in a reasoning test with accuracy feedback

Nadin Beckmann[1]

Jens F. Beckmann[2]

Julian G. Elliott[3]

Corresponding author:

[1]Australian School of Business

Accelerated Learning Laboratory

University of New South Wales

Sydney, NSW 2052, Australia

Email: nadinb@agsm.edu.au

Tel: +61 2 9931-9187

Fax: +61 2 9931-9199

[2]School of Psychology

University of Sydney

Brennan MacCallum Building (A18)

Sydney, NSW 2006, Australia

Email: jensb@psych.usyd.edu.au

[3]School of Education

Durham University

Leazes Road

Durham DH1 1TA, United Kingdom

Email: joe.elliott@durham.ac.uk

Self-confidence and performance goal orientation interactively predict performance in a reasoning test with accuracy feedback

1 Introduction

Performance feedback is widely considered to be a prominent agent of learning processes or at least of improved performance. However, research has shown that, counter-intuitively, providing feedback does not always improve task performance. Meta-analytic studies have provided evidence that feedback effects on performance are often quite variable; on some occasions, feedback improves performance, in others, no effect can be found, and yet in other instances, feedback can negatively influence performance (Bangert-Drowns, Kulik, Kulik & Morgan, 1991; Kluger & DeNisi, 1996).

Relatively little is known about feedback effects in psychometric assessment settings, although the few studies published reveal a similarly mixed picture. Item feedback has been found to enhance performance in cognitive ability and achievement tests (e.g., Betz, 1977; Birenbaum & Tatsouka, 1987; Dash & Rath, 1984; Dillon, 1981; Bethge, Carlson & Wiedl, 1982; Carlson & Wiedl, 1979), have no effect on test performance (e.g., Delgado & Prieto, 2003; Roos, Wise & Plake, 1997; Stankov & Crawford, 1997; Wise, Plake, Pozehl, Boettcher-Barnes, Lukin, 1989), and, in other studies, negatively affect test performance (e.g., Delgado & Prieto, 2003; Strang & Rust, 1973; Wise, Plake, Eastman Boetcher & Lukin, 1986).

The provision of feedback is *the* central feature of dynamic testing (for an overview see Lidz & Elliott, 2000). Dynamic tests typically consist of items similar to those usually utilized in intelligence tests (e.g., reasoning problems) ordered in increasing complexity. However, dynamic tests differ procedurally from traditional measures in that the examinee is provided with feedback about their responses to items. Because this offers an opportunity to learn during the test session, it is argued that performance on such measures offer more valid indication of cognitive abilities (e.g., learning potential) than test scores obtained from traditional feedback-free intelligence tests (Grigorenko & Sternberg, 1998; Guthke & Beckmann, 2003; Guthke & Stein, 1996).

The simplest and, hence, most common form of feedback provided in dynamic tests and other cognitive ability tests concerns whether a response to a test item is accurate or not (in dynamic tests, e.g., Guthke & Beckmann, 2000; other intelligence tests, e.g., Stankov & Crawford, 1997; and adaptive achievement tests, e.g., Tonidandal, Quinones, & Adams, 2002). However, the mechanisms by which this very simple form of feedback affects test performance are far from understood. On a theoretical level Carlson and Wiedl (1992, 2000) discuss three functions of feedback in cognitive ability testing: modification, compensation and inhibition. With regard to simple accuracy feedback it could be argued that it can *modify* knowledge related to problems by informing test takers about the appropriateness of the problem solving strategies they have employed; it might *compensate* for factors hindering performance by relocating the test taker's focus of attention to the task, and it might *inhibit* performance-reducing behaviours by motivating test takers to work harder. We argue however, accuracy feedback could also undermine the confidence of test takers and thus, negatively affect their test performance.

Heterogeneous result patterns as observed in feedback research would seem to suggest the involvement of moderating variables. Those can lie in the situation (e.g., level of elaborateness of feedback) or in the person (e.g., individual differences variables). Focussing on one type of feedback – accuracy feedback – the current study tests this assumption by analysing the potentially moderating effects of two person variables; goal orientation and self-confidence.

Goal theory describes an important motivational mechanism that may be affected by feedback. Those pursuing *learning goals* tend to be concerned with increasing their competence by acquiring new skills and mastering new tasks. This focus has been shown to be beneficial to performance across a range of tasks (for a meta-analytic review of experimental studies see Utman, 1997). As a learning goal orientation is associated with the belief that performance can be increased through higher levels of effort (Dweck, 1986, 1999), feedback may be seen as providing helpful information in this respect, helping those individuals to adopt and maintain effective problem solving strategies (Cron, Slocum, & VandeWalle, 2005; Davis, Carson, Ammeter, & Treadway, 2005; VandeWalle, Cron, & Slocum, 2001).

Individuals pursuing *performance goals*, on the other hand, may have a disadvantage over their peers in achievement settings when confronted with failure. Performance goal oriented individuals tend to be concerned with demonstrating and validating their competence by seeking positive judgements and/or avoiding negative judgements. Their primary focus is on performance outcomes, such as their grades or test results, and how these are being perceived by others (e.g., teachers or peers). This has been shown to be detrimental to task performance, where individuals have low confidence in their current ability (Dweck, 1986, 1999; Elliott & Dweck, 1988). As performance goal orientation is often related to a belief that performance is based on innate and therefore fixed ability (Dweck, 1986, 1999), individuals pursuing such goals might be more likely to interpret feedback primarily as an evaluation of their ability. Failure feedback, therefore, may undermine subsequent performance (Cron, et al., 2005; Davis, et al., 2005; VandeWalle, et al., 2001).

When constraining the focus on accuracy feedback the result pattern homogenises to small and zero performance effects (Carlson & Wiedl, 1979; Dash & Rath, 1984; Dillon, 1981; Kluger & DeNisi, 1996). This indicates that the level of elaborateness of feedback acts as a situation-related moderator variable. Our main goal in this study is to examine the involvement of person-related moderating variables that may result in differential effects, which, ultimately, may threaten the validity of cognitive ability tests employing accuracy feedback imprudently. In terms of a baseline, from a *general* perspective we hypothesize:

> H1: On average, accuracy feedback provided after each item response will have no, or negligible, effects on performance on a cognitive ability test.

Following Dweck and colleagues (1986, 1988, 1999) who have demonstrated that a performance goal orientation results in maladaptive affective-cognitive and behavioural response patterns only when it occurs in combination with low self-confidence we further hypothesize from a *differential* perspective:

> H2: A performance goal orientation paired with low levels of self-confidence will negatively affect performance in a cognitive ability test when accuracy feedback is provided after each item response.

2 Method

2.1 Participants

The study involved 105 students of a mainstream secondary school (aged 13 to 15, 43% female) in the North East of England, UK. Students participated voluntarily. Only two students chose to withdraw from the study.

2.2 Measures

2.2.1 Experimental Task

The main dependent variable was the total number of correct responses from a set of twelve number-based reasoning problems presented in increasing complexity. Each of the computer-based presented items consisted of seven numbers. The task was to enter the eighth number that continued the series according to an underlying rule. On average, it took about 20 minutes for the students to tackle all of the problems. Internal consistency was high for the two parallel versions used ($\alpha = .81$).

Two reasoning test conditions were employed, a feedback and a feedback-free condition. In the feedback condition test takers received accuracy feedback (a visual message worded “correct” or “incorrect”) as a consequence of their response to each item. In the feedback-free condition no feedback was provided.

2.2.2 Intelligence Test

We had access to the students' psychometric intelligence test scores routinely collected by their school, which were based on either the Middle Years Information System for eighth graders (MidYIS, Durham University), or the Cognitive Abilities Test (CAT, Thorndike & Hagen, 1993) for ninth graders, respectively. The MidYIS contains tasks that assess vocabulary, mathematical skills, information processing speed, spatial abilities, and reasoning abilities. The CAT assesses reasoning ability in three domains (verbal, non-verbal, numeric). Standardized scores were combined into one variable representing the psychometric intelligence of the participants based on either the MidYIS or the CAT (IQ-scores).

2.2.3 Self-Report Measures

Paper and pencil questionnaires were administered to assess goal orientation and self-confidence. Based on validated inventories (Motivational Orientation Scales, Duda & Nicholls, 1992, 16 items; Self-confidence and Goal Orientation, Dweck, 1999, 3 and 4 items, respectively) we developed state-oriented measures focussing on the reasoning problems at hand. Three task-specific goal orientation items; one performance goal item, one learning goal item, and one item that obliged the participant to make a forced choice between performance and learning goal orientation were utilised. Confidence in test performance was measured with one item (see Appendix A). The answer format for all items was a visual analogue scale. Participants responded to each statement by placing a cross along a seven-centimetre line with the polar ends labelled "strongly agree" to statement A on one side and "strongly agree" to statement B on the other – the two statements being in direct opposition to each other (see Appendix A).

In order to check whether the modified and shortened versions still refer to the same constructs as their original trait-oriented counter-parts the complete scales in their original versions were also administered (see Table 1 for correlations).

2.3 Design

A semi-balanced repeated measurement design was employed. To enable intra-individual comparisons between performance in the feedback and the feedback-free conditions each student was presented with two parallel versions of the computerized reasoning test, undertaking one on each of two occasions. Students were randomly assigned to either one of two experimental groups or to the control group. To control for sequence effects, experimental group 1 received feedback only in the second test session (denoted as F–F+); and experimental group 2 received feedback solely in the first test session (denoted as F+F–). Students in the control group (F–F–) received no feedback in either test session. To identify differential effects of feedback, we were particularly interested in experimental group 1 (F–F+), and, for this reason, a greater proportion of students was assigned to this group ($N_{F-F+} = 50$, $N_{F+F-} = 28$, $N_{F-F-} = 27$). The three groups did not differ in terms of their psychometric intelligence (mean $IQ_{F-F+} = 104.8$, $SD = 12.7$; mean $IQ_{F+F-} = 106.5$, $SD = 11.5$; $IQ_{F-F-} = 100.2$, $SD = 10.8$; $F_{3, 103} = 0.73$, $p = .49$).

2.4 Procedure

The study comprised three sessions. Immediately prior to the administration of the reasoning tests in session 1, students were presented with two example items (Appendix B).

While presented with this information on the computer screen, they were asked to answer the task-specific questions on paper about how they thought and felt about taking the test (see Appendix A). Students were then asked to tackle the computerised number series problems. Session 2 took place one or two days afterwards. On this occasion, students were given the trait-oriented questionnaires containing items referring to their trait (i.e., their general, task-unspecific) goal orientation and level of academic self-confidence. Session 3 took place one or two days after session 2. As for session 1, the computerised reasoning tests, together with task-specific questions regarding self-confidence and goal orientation, were administered.

2.5 Data Analysis

To test hypothesis 1, contrast analyses were carried out. The five contrasts tested were: a) the inter-individual effect of feedback on performance (two contrasts), b) the practice effect on performance (one contrast), and c) the intra-individual effect of feedback on performance, i.e. the interaction between group membership and test session on performance (two contrasts).

To test hypothesis 2, three moderated multiple regression analyses (MMR, e.g., Aguinis, 2004; Cohen, Cohen, West, & Aiken, 2003) of reasoning test performance were carried out. In all three MMR analyses, psychometric intelligence was controlled for by including intelligence test performance as a covariate in the analyses. In addition to intelligence test performance, task-specific performance goal orientation and task-specific self-confidence, were included as independent variables in the analyses. All independent variables were mean centred.

3 Results

Table 1 shows the means, standard deviations and item-intercorrelations for the study variables based on the total sample across all experimental conditions.

Insert Table 1 about here

3.1 Effect of Accuracy Feedback on Performance

Table 2 displays the means and standard deviations for the dependent variable, reasoning test performance, in each experimental condition.

Insert Table 2 about here

In order to analyse whether the provision of accuracy feedback affected students' reasoning test performance (Hypothesis 1) we compared average performance levels under feedback and feedback-free conditions taking a between- and within-subject perspective (see Table 3). Between-subject contrast analyses as well as within-subject contrast analyses showed that, as expected, there was no significant feedback effect on performance in either test session. No practice effect was observed. These findings support hypothesis 1 stating that accuracy feedback would, on average, have no effect on performance in a reasoning test.

Insert Table 3 about here

With regard to hypothesis 2 we examined whether individual differences in task-specific goal orientation and self-confidence explain performance differences under feedback conditions compared with feedback-free conditions.

3.2 Interaction Effect of Self-confidence and Performance Goal Orientation

3.2.1 Feedback Condition

Under feedback conditions a significant interaction effect between task-specific self-confidence and task-specific performance goal orientation on reasoning test performance can be observed ($R^2_{adj} = .50$, $F_{4,49} = 13.12$, $p < .001$, $\beta_{\text{product}} = -.21$, $p = .05$). In addition, self-confidence significantly predicted performance ($\beta = .26$, $p < .05$) above and beyond any of the other tested effects. The interaction effect explained an additional 4% of the variance in the dependent variable.[1]

The first panel in Figure 1 depicts the significant interaction effect under feedback conditions (experimental group 1, F–F+, test 2, $N = 50$), which will be interpreted together with the results obtained for the feedback-free condition.

Insert Figure 1 about here

3.2.2 Feedback-free Condition

As expected, when analysing the feedback-free reasoning test performance of the same group (F–F+; test 1) no interaction between task-specific self-confidence and task-specific goal orientation was detected ($R^2_{adj} = .54$, $F_{4,49} = 15.38$, $p < .001$, $\beta_{\text{product}} = .14$, ns). The regression coefficients for both conditions, feedback and feedback-free, differ significantly (CI for $\alpha = 5\%$: .06 to .59, Cohen et al., 2003). Given the statistical test conditions (i.e., sample size, significance threshold) the existence of a moderate to strong interaction effect ($f^2 \geq .17$) of self-confidence and performance goal orientation on *feedback-free* reasoning test performance can be ruled out with sufficient statistical power ($1-\beta \geq .80$). Self-confidence again predicted performance significantly ($\beta = .42$, $p < .001$) above and beyond any of the other tested effects.

For comparison purposes, the second panel in Figure 1 shows the non-significant interaction effect in the same sample under *feedback-free* conditions (experimental group 1, F–F+, test 1). Self-confidence prior to testing was positively related to test performance, irrespective of the examinee's goal orientation and psychometric intelligence, as indicated by the significant main effect depicted in both panels of Figure 1. In addition, more confident examinees with low performance goal orientation gained from feedback (intra-individually up to 1.3 tasks; compare panel 1 and panel 2 in Figure 1). However, rather unexpectedly, examinees low in self-confidence also benefited from feedback (intra-individually up to 2.4 tasks), if they scored highly in respect of performance goal orientation.

3.2.3 Practice Effect

The absence of an interaction effect of self-confidence and goal orientation on reasoning test performance in session 2 (no feedback but parallel test session) in the control (F–F–) and experimental group 2 (F+F–) combined[2] (R^2_{adj} = .23, $F_{4,53}$ = 4.99, p < .05, β_{product} = .01, ns) indicates that the moderator effect observed for experimental group 1 (F–F+) under feedback conditions is not a result of a practice or parallel test effect.

4 Discussion

This study sought to investigate individual differences in responding to item specific accuracy feedback in psychometric ability testing. On average, implementing accuracy feedback in the reasoning tests appeared neither to improve nor undermine test takers' performance (Hypothesis 1). This is in line with findings reported in those few studies on accuracy feedback in cognitive ability testing that are currently available (Delgado & Prieto, 2003; Stankov & Crawford, 1997; Wise et al., 1989). However, differential effects were identified when considering moderator variables. Examinees low in self-confidence benefited from accuracy feedback if they pursued performance goals (Hypothesis 2). For this group a competitive motivational orientation towards testing seemed to have a compensatory effect enabling them to reach a performance level similar to those with high self-confidence.

This finding appears meaningful for three reasons: a) The observed differential effect cannot be attributed to differences in intellectual ability as we controlled for psychometric intelligence in our analyses; b) Contrasting feedback and feedback-free test conditions allowed us to conclude that the likelihood of such an effect is considerably smaller in test conditions where no feedback is provided; and, c) the utilisation of a combined within- and between-subject design enabled us to demonstrate that there are meaningful individual differences in responding to accuracy feedback that would have been masked in a more general study of feedback effects.

In contrast to our initial expectation with regard to hypothesis 2, test takers who expressed low confidence in their ability and also held performance goals benefited most from accuracy feedback. We discuss three reasons why this might be the case. First, performance goals as measured in our study by such items as, *"I think I will have done well on this test if I get more correct answers that other students"*, have been described as approach performance goals as opposed to avoid performance goals (e.g., avoiding displaying potentially low performance, Elliott & Church, 1997). In two more recent studies on self-regulatory responses to performance feedback, negative performance effects (VandeWalle, et al., 2001) and negative emotional reactions to negative feedback (Cron, et al., 2005) have been traced back to the avoidance rather than to the approach aspect of performance goal orientation. Second, test-takers pursuing performance goals may have had an advantage over their peers in terms of a person-situation fit. Test conditions with feedback can be perceived as strongly evaluative situations, which could accommodate a competitive motivational orientation. Third, in contrast to most previous research, in the current study participants received valid feedback that represented their true performance. While the moderate overall test difficulty ($\overline{p}_{\text{versionA}} = .60$, $\overline{p}_{\text{versionB}} = .62$) observed in the present sample suggests that, on average, all examinees experienced a degree of failure they also all received a certain amount of positive feedback.

Our findings suggest that accuracy feedback induces systematic variance in test performance independently of cognitive ability. This throws into question the validity of cognitive ability tests that employ feedback without regard to the interplay of non-intellective factors. Future research should investigate whether the reported effects generalise to more complex forms of feedback than are employed in this study.

The inclusion of feedback in cognitive ability testing needs to be considered carefully. It is recognised that underperformance can occur when existing competencies are not fully utilised (e.g., because of non-intellective factors such as anxiety), but, where the proper procedures are adhered to, it is theoretically impossible to achieve over-performance (i.e., test performance cannot be systematically better than the individual's actual ability level permits). Hence, the higher performance score, under feedback conditions, of test takers reporting low pre-test confidence and performance goals could be interpreted as a more meaningful indicator of their "true" ability. It is conceivable, therefore, that feedback free test conditions may place certain subgroups (e.g., those with low pre-test confidence and performance goals) at a disadvantage. On the other hand, the fact that the observed performance increase can be explained by non-intellective factors leads us to question whether performance scores in feedback conditions should indeed be interpreted as "pure" measures of cognitive abilities.[3]

Carlson and Wiedl (2000) conceptualise test performance as a function of three variables: the type of test items, the testing procedure (e.g., provision of feedback), and the individual. Importantly, individual differences in ability and personality and procedures used in cognitive ability tests interact to influence test performance. Thus, a testing procedure that optimises performance for one group might not for another (Carlson & Wiedl, 1992; Wiedl & Carlson, 1985). The data provided in the current study supports this view. Future research should put more emphasis on investigating the interplay between person variables and presentation modes of cognitive ability tests. Certainly, until further research clarifies the role of individual differences such as those examined in the present study, test developers, test administrators and practitioners need to be alert to the complexities of introducing feedback into cognitive testing.

5 Footnotes

[1] The same pattern of results emerges when employing a bipolar goal orientation approach *"What is important to me is getting a higher score on the test than everyone else"* vs. *"What is important to me is learning how to solve the problems"* (R^2_{adj} = .50, $F_{4,49}$ = 13.10, p < .001, ΔR^2 = .04, $F_{1,45}$ = 4.28, p < .05). Equivalent analyses using task-specific *learning* goal orientation as the independent variable instead of task-specific performance goal orientation revealed no interaction effect.

[2] The combination of both groups results in a sample size of N = 54 and hence in a similar statistical power as for the analyses conducted for experimental group 1 (F–F+).

[3] Here, we should not forget that the intellectual ability scores that served as a covariate in our analyses were obtained under feedback-free conditions (the intelligence test batteries used by the school do not provide feedback). Therefore, we basically controlled for intellectual ability under feedback-free conditions in our analyses.

6 References

Aguinis, H. (2004). *Regression analysis for categorical moderators*. New York, NY: Guilford Press.

Bangert-Drowns, R. L., Kulik, C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*, 213-238.

Bethge, H. J., Carlson, J. S., & Wiedl, K. H. (1982). The effects of dynamic assessment procedures on Raven Matrices performance, visual search behavior, test anxiety and test orientation. *Intelligence, 6*, 89-97.

Betz, N. E. (1977). Effects of immediate knowledge of results and adaptive testing on ability test performance. *Applied Psychological Measurement, 1*, 259-266.

Birenbaum, M., & Tatsuoka, K. (1987). Effects of "online" feedback on the seriousness of subsequent errors. *Journal of Educational Measurement, 24*, 145-155.

Bird, K. D. (2004). *Analysis of variance via confidence intervals*. London, UK: Sage Publications.

Carlson, J. S., & Wiedl, K. H. (1979). Toward a differential testing approach: Testing-the-limits employing the Raven matrices. *Intelligence, 3*, 323-344.

Carlson, J. S., & Wiedl, K. H. (1992). Principles of dynamic assessment: The application of a specific model. *Learning & Individual Differences, 4*, 153-166.

Carlson, J., & Wiedl, K. H. (2000). The validity of dynamic assessment. In C. S. Lidz & J. G. Elliott (Eds.), *Dynamic Assessment: Prevailing models and applications* (Vol. 6, pp. 681-712). New York, NY: Elsevier.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioural sciences* (3 ed.). Mahwah, NJ: Erlbaum.

Cron, W. L., Slocum, J. W., Jr., VandeWalle, D., & Fu, Q. (2005). The role of goal orientation on negative emotions and goal setting when initial performance falls short of one's performance goal. *Human Performance, 18*, 55-80.

Dash, A. S., & Rath, S. (1984). Dynamic testing approaches to the study of cognitive development: Effects on progressive matrices scores. *Psycho-Lingua, 14*, 35-46.

Davis, W. D., Carson, C. M., Ammeter, A. P., & Treadway, D. C. (2005). The interactive effects of goal orientation and feedback specificity on task performance. *Human Performance, 18*, 409-426.

Delgado, A. R., & Prieto, G. (2003). The effect of item feedback on multiple-choice test responses. *British Journal of Psychology, 94*, 73-85.

Dillon, R. F. (1981). Analogical reasoning under different methods of test administration. *Applied Psychological Measurement, 5*, 341-347.

Duda, J. L., & Nicholls, J. G. (1992). Dimensions of achievement motivation in schoolwork and sport. *Journal of Educational Psychology, 84*, 290-299.

Dweck, C. S. (1986). Motivational processes affecting learning. Special Issue: Psychological science and education. *American Psychologist, 41*, 1040-1048.

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development*. Philadelphia: Psychology Press.

Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology, 54*, 5-12.

Grigorenko, E., & Sternberg, R. (1998). Dynamic testing. *Psychological Bulletin, 124*, 75-111.

Guthke, J., & Beckmann, J. F. (2000). The learning test concept and its application in practice. In C. S. Lidz & J. G. Elliott (Eds.), *Dynamic Assessment: Prevailing models and applications* (Vol. 6, pp. 17-69). Oxford, UK: Elsevier.

Guthke, J., & Beckmann, J. F. (2003). Dynamic Assessment with Diagnostic Programs. In R. J. Sternberg, J. Lautrey & T. I. Lubart (Eds.), *Models of Intelligence: International Perspectives.* Washington, DC: American Psychological Association.

Guthke, J., & Stein, H. (1996). Are learning tests the better version of intelligence tests? *European Journal of Psychological Assessment, 12*, 1-13.

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.

Lidz, C., & Elliott, J. G. (Eds.). (2000). *Dynamic assessment: Prevailing models and applications*. New York, NY: Elsevier Science.

Roos, L. L., Wise, S. L., & Plake, B. S. (1997). The role of item feedback in self-adapted testing. *Educational and Psychological Measurement, 57*, 85-98.

Stankov, L., & Crawford, J. D. (1997). Self-confidence and performance on tests of cognitive abilities. *Intelligence, 25*, 93-109.

Strang, H. R., & Rust, J. O. (1973). The effects of immediate knowledge of results and task definition on multiple-choice answering. *Journal of Experimental Education, 42*, 77-80.

Thorndike, R. L., & Hagen, E. P. (1993). *Cognitive Abilities Test*. Itasca, IL: Riverside Publishing.

Tonidandal, S., Quinones, M. A., & Adams, A. A. (2002). Computer adaptive testing: The impact of test characteristics on perceived performance and test takers' reactions. *Journal of Applied Psychology, 87*, 320-332.

Utman, C. H. (1997). Performance effects of motivational state: A Meta-Analysis. *Personality and Social Psychology Review, 1*, 170-182.

VandeWalle, D., Cron, W. L., & Slocum, J. W. (2001). The role of goal orientation following performance feedback. *Journal of Applied Psychology, 86*, 629-640.

VandeWalle, D., & Cummings, L. L. (1997). A test of the influence of goal orientation on the feedback seeking process. *Journal of Applied Psychology, 82*, 390-400.

Wise, S. L., Plake, B. S., Eastman, L. A., Boettcher, L. L., & Lukin, L. (1986). The effects of item feedback and examinee control on test performance and anxiety in a computer-administered test. *Computers in Human Behavior, 2*, 21-29.

Wise, S. L., Plake, B. S., Pozehl, B. J., Boettcher-Barnes, L. B., & Lukin, L. (1989). Providing item feedback in computer-based tests: Effects of initial success and failure. *Educational and Psychological Measurement, 49*, 479-486.

7 Figure Captions

Figure 1. Numeric reasoning test performance in experimental group 1 (*N* = 50) with group specific one standard deviation below (“low”) and above (“high”) average scores for the predictors “performance goal orientation” (item 2 in Appendix A) and “self-confidence” (item 1 in Appendix A), and average intellectual ability under a) feedback conditions (left panel) and b) feedback-free conditions (right panel). Arrows indicate the intra-individual performance improvement under the feedback condition.

Figure 1

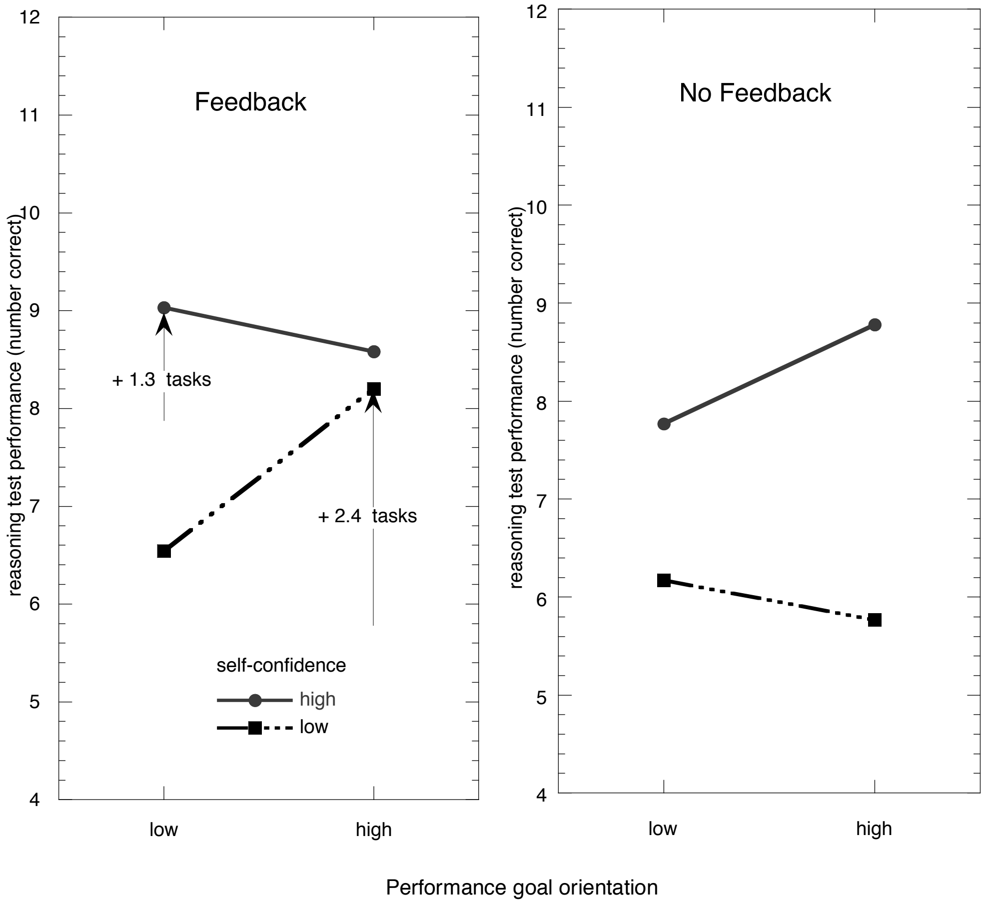

Figure 1

Table 1

Table 1: Descriptive Statistics for Study Variables in the Complete Sample Across Experimental Conditions (N = 104 to 105)

Variables	*Mean*	*SD*	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
Test 1 and state measures															
1. Performance (number correct)	7.54	2.64	(.81)												
2. Pre-test task-specific performance goal	4.12	1.85	.14												
3. Pre-test task-specific learning goal	4.72	1.85	-.02	-.02											
4. Pre-test task-specific performance vs. learning goal	2.23	1.99	-.003	.29**	-.27**										
5. Pre-test task-specific self-confidence	4.84	1.52	.48**	.21*	-.04	.07									
Test 2 and state measures															
6. Performance (number correct)	7.64	2.80	.77**	.15	.04	.12	.42**	(.81)							
7. Pre-test task-specific performance goal	4.40	1.59	.15	.20*	-.16	.24*	.27**	.16							
8. Pre-test task-specific learning goal	4.53	2.02	-.12	-.12	.52**	-.36**	.16	-.15	-.14						
9. Pre-test task-specific performance vs. learning goal	3.00	2.16	.07	.17	.17	.48**	.17	.07	.36**	-.33**					
10. Pre-test task-specific self-confidence	4.57	1.79	.44**	.11	.15	.13	.51**	.39**	.16	.08	.11				
Trait measures															
11. Performance goal orientation	0.00	1.00	.14	.24*	-.17	.30**	.35**	.19	.52**	-.24*	.51**	.26**	(.90)		
12. Learning goal orientation	0.02	1.01	.16	.07	.35**	-.31*	.21*	.07	-.06	.28**	-.11	.24*	-.04	(.85)	
13. Self-confidence	4.43	1.65	.34**	.28*	.20*	.01	.33*	.27**	.02	.05	.02	.50**	.18	.38**	(.70)
14. Intelligence test performance	106.13	11.79	.58**	.03	-.09	-.05	.18	.51**	.12	-.08	.03	.13	.01	.14	.25*

Note: For performance goal orientation (11.) and learning goal orientation (12.) factor scores were calculated based on a larger sample (N = 419) including data from other studies conducted by the authors that employed the respective scales. **p < .01, *p < .05; Coefficients in brackets represent Cronbach's α for the respective scales

Table 2

Table 2: Descriptives of Reasoning Test Performance (Amount Correct) in Each of the Experimental Conditions

	N	*Test 1*			*Test 2*		
		feedback	*Mean*	*SD*	*feedback*	*Mean*	*SD*
Experimental Group 1 (F–F+)	50	no	7.20	2.76	yes	7.90	2.73
Experimental Group 2 (F+F–)	28	yes	7.68	2.31	no	6.96	2.50
Control Group (F–F–)	27	no	8.04	2.74	no	7.85	3.21

Table 3

Table 3: Contrasts Coefficients and Confidence Intervals for all Tested Experimental Effects

Effect		Contrast	F (df)	CI limits Lower	CI limits Upper
1)	Between-subject effect of feedback in session 1	Experimental group 2 vs. control group (at time 1)	0.25 (1,102)	-0.67	0.40
2)	Between-subject effect of feedback in session 2	Experimental group 1 vs. control group and experimental group 2 combined (at time 2)[4]	0.81 (1,102)	-0.21	0.56
3)	Practice effect (repeated measurement with parallel tests)	Test session 1 vs. test session 2	0.14 (1,102)	-0.11	0.16
4)	Within-subject effect of feedback in session 2	Experimental group 1 vs. control group by test session	4.43 (1,102)	-0.68	0.03
5)	Within-subject effect of feedback in session 1	Experimental group 2 vs. control group by test session	1.24 (1,102)	-0.20	0.59

Note: Confidence interval (CI) limits are Bonferroni adjusted and standardized, scaled in sample SD units (Bird, 2004); all CIs indicate non-significance; [4]As there were no significant differences in reasoning test performance observed between experimental group 2 (F+F–) and control group (F–F–), either at time 1 or time 2, for test session 2 (no feedback) both groups were analysed simultaneously to increase statistical power.

8 Appendix A

Questions to assess task-specific self-confidence (item 1) and goal orientation (performance goal, item 2; learning goal, item 3; and forced choice between performance and learning goal, item 4) prior to the reasoning test

1. I feel pretty confident that I shall be able to solve most of the problems. vs. I do *not* feel confident that I shall be able to solve most of the problems.
2. I think I will have done well on this test if I get more correct answers than other students. vs. I do *not* think that I will have done well unless I get more correct answers than other students.
3. Even if I get most of the problems wrong, I will feel that I have done *well* if I learn something interesting. vs. If I get most of the problems wrong, I will feel that I have done *poorly* even if I learn something interesting.
4. What is important to me is getting a *higher score* on the test than everyone else. vs. What is important to me is *learning* how to solve the problems.

The answer format was an analogue scale from "strongly agree" to statement A to "strongly agree" to statement B.

9 Appendix B

Computerized instruction presented while answering the task-specific goal orientation and self-confidence items prior to the reasoning test

All the problems that you will receive will be similar to these:

The first one is a pretty easy problem.

33 28 26 21 19 14 12 ?

Which number follows?

The second problem is an example for the most difficult ones in the test.

4 6 18 21 84 88 440 ?

Which number follows?

All the questions in the questionnaire concern the problems presented on the computer.
Please start now to fill in the questionnaire.