Communications of the IIMA

Volume 6 | Issue 3 Article 14

2006

Evaluation of User Acceptance for Web-based Aptitude Tests

Michael Amberg
University of Erlangen-Nuremberg

Sonja Fischer
University of Erlangen-Nuremberg

Manuela Schroder
University of Erlangen-Nuremberg

Evaluation of User Acceptance for Web-based Aptitude Tests

Michael Amberg
University of Erlangen-Nuremberg, Lange Gasse 20, 90403 Nuremberg, Germany,
+49-(0)911-5302-801, +49-(0)911-5302-860, amberg@wiso.uni-erlangen.de

Sonja Fischer
University of Erlangen-Nuremberg, Lange Gasse 20, 90403 Nuremberg, Germany,
+49-(0)911-5302-876, +49-(0)911-5302-860, sonja.fischer@wiso.uni-erlangen.de

Manuela Schröder
University of Erlangen-Nuremberg, Lange Gasse 20, 90403 Nuremberg, Germany,
+49-(0)911-5302-863, +49-(0)911-5302-860, manuela.schroeder@wiso.uni-erlangen.de

ABSTRACT

Web-based aptitude tests, which are a special category of aptitude tests, can be used for rather standardized test methods and for a large amount of users. The characteristics of web-based aptitude tests can have an impact on the test results and the user acceptance. The aim of our research is to develop a method for the evaluation of the user acceptance for web-based aptitude tests. Therefore, we used the DART-approach with the dimension (Perceived) Usefulness, (Perceived) Ease of Use, (Perceived) Network Effects and (Perceived) Costs as the theoretical basis, identified important acceptance indicators, developed a questionnaire and conducted a survey. Afterwards, we proved the reliability and conducted a factor analysis. The results point out that some of the defined acceptance indicators should be revised. Additionally, the factor analysis shows that a combination of two dimensions (Perceived) Usefulness and (Perceived) Network Effects is useful especially with regard to web-based aptitude tests. Finally, we conducted a univariate analysis to evaluate the user acceptance of a web-based aptitude test. The visualised result on the basis of a DART-chart clearly shows that the interviewees evaluated the indicators very differently. There are fields, where the aptitude test fulfils the expectations, and fields, which can be improved.

Key words: web-based aptitude test, user acceptance, DART

INTRODUCTION

Web-based aptitude tests are getting more and more important in different fields of application. Especially, the study choice and the applicant selection can be supported by these tests. Generally, web-based aptitude tests determine the competence of a person for school, study, special jobs or tasks. Furthermore, they are characterised by low costs, high effectiveness, overall availability and an easy distribution of the test questions and the test results. Predominantly, they can be used for rather standardized test methods and for a large amount of users.

At universities in Germany, there are only few web-based aptitude tests to support applicants to choose their field of study. There are many endeavours of different German universities to develop web-based aptitude test. The objectives are to assess a large amount of high-school graduates with relatively low costs and high effectiveness and to find the applicants, who likely pass the university. At the moment, there are no consolidated findings about the designing of web-based aptitude test and the influence of the user interface on the user acceptance and the test results.

Hornke and Zimmerhofer (2005), who were involved in the development of the web-based aptitude test at the university (RWTH Aachen), tried to measure the user acceptance. Therefore, they used the following criteria:

- A general evaluation based on a scale from 1 (very good) to 6 (bad)
- The user's opinion about the information content for the study choice

- The rate how often the user would recommend the web-based aptitude test

The survey does not include the analysis of factors or indicators, which influence the user acceptance of web-based aptitude tests. However, these findings could help to improve the user acceptance. A high user acceptance is important, since then test results tend to be taken more seriously. One target of our research is to identify indicators, which have an impact on the user acceptance for web-based aptitude tests and can be directly influenced by the developers of web-based aptitude tests. According to the Dynamic Acceptance Model for the Re-evaluation of Technologies (DART), we identified acceptance indicators, developed a questionnaire and conducted a survey, which results show improvement opportunities. On the basis of our results, we finally analysed and discussed the user acceptance of the selected web-based aptitude test.

THEORETICAL BACKGROUND

In general, acceptance is defined as an antagonism to the term refusal and means the positive decision to use an innovation. The acceptance research has its origin in both, industrial and business science. While industrial science focuses on the conditions of user friendly technologies and techniques, the business science discusses acceptance in a variety of disciplines (e. g., marketing, organisation and production theory as well as information management) (Amberg, Hirschmeier, Wehrmann, 2004).

During the last years several, theories and models have been developed. Each of these models determines different factors to explain user acceptance. The following table shows an overview of the most widely used acceptance models.

	Key influencing factors	Short summary
Technology Acceptance Model (TAM) (Davis, 1986)	• (perceived) usefulness • (perceived) ease of use	Benefit and effort as basic distinction for an influence factors on the acceptance decision of innovative technologies.
Technology Task Fit Model (TTFM) (Goodhue, Thompson 1995)	• Technology • Task • Individual	Task oriented approach to address the acceptance of IT systems.
(Degenhardt, 1986)	• Task • System • User Characteristics	Acceptance of communication services at the example of interactive videotext (BTX).
Dynamic Acceptance Model (Kollmann, 1998)	• attitude level • action level • utilisation level	Process related interpretation of the user acceptance.
Dynamic Acceptance Model for the Reevaluation of Technologies (DART) (Amberg, Hirschmeier, Wehrmann, 2004)	• (perceived) usefulness • (perceived) ease of use • (perceived) networkeffects • (perceived) costs	Continuous reevaluation of the user acceptance based on a meta structure for the systematic determination of a well balanced and complete set of individually measurable influencing factors.

Table 1: Overview of common acceptance models and their key influencing factors (according to Amberg et al., 2004)

A valuable contribution in this context is an acceptance model, namely **DART-approach**, which was introduced by (Amberg et al., 2004). DART is an instrument designed for the analysis and evaluation of the user acceptance of products or services. The fundamental design criteria are according to Amberg et al. (2004):

- Use as a permanent controlling instrument
- Balanced consideration of relevant influencing factors

- Applicability during the whole development/implementation process
- Adaptability to individual requirements of the research item

These design criteria are useful to integrate user acceptance analysis into the development and evaluation of products or services. In the following, we describe the architecture of DART (Amberg et al., 2004). DART, which is based on the idea of the balanced scorecard (BSC) (cf. Kaplan and Norten, 1996), helps to define a balanced set of measurable acceptance indicators for the evaluation of the user acceptance. The DART-approach uses a meta-structure, which consists of the following complementary and orthogonal categories:

- **Benefits** and **Efforts** comprise all positive and negative facets of the products or services under examination.
- **Products** or **Services** (e. g., Internet Applications) and Contextual Conditions of Products or Services include basic sociocultural and economic conditions, which also have a considerable impact on user acceptance.

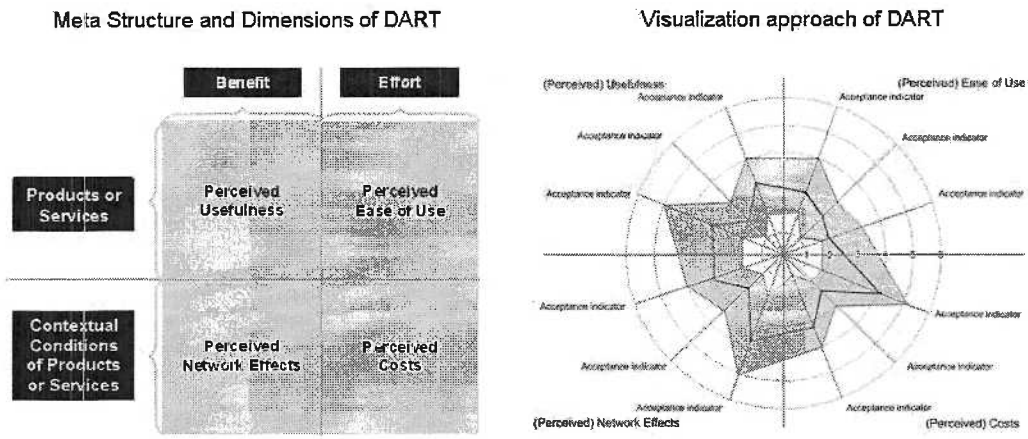

Figure 1: The DART-approach (Amberg et al., 2004)

These categories lead to four dimensions (figure 1):

- **(Perceived) Usefulness**: The dimension build by the categories Benefits and Products or Services describes the perceived usefulness of a product or service. Indicators measuring this dimension might be perceived information quality and quantity or conformity of expectations.
- **(Perceived) Ease of Use**: The dimension characterized by the categories of Products or Services and Efforts can be identified with the perceived ease of use, the degree to which a person believes that using a particular service would be free of effort. Indicators measuring this dimension are for example the ease of configuration or first log-in, overall handling and menu navigation.
- **(Perceived) Network Effects**: The categories Benefits and Contextual Conditions of Products or Services lead to the dimension of perceived network effects. The dimension considers the contextual conditions of a product or a service, whose acceptance highly depends on the economical, social and technological perspectives (Galletta and Malhotra, 1999).
- **(Perceived) Costs**: This dimension is formed by the categories of Efforts and General Conditions of Products or Services. This dimension describes the monetary and non-monetary effort. Costs transparency or data security, which are not in correlation with the application, are considered as appropriate indicators.

These four dimensions emphasize the evaluation of a product or a service from the end user's subjective point of view, because of the explicit consideration of the user's perception. On the basis of these dimensions, it is possible to define specific acceptance indicators for the empirical evaluation of the user acceptance respectively.

The **visualization approach** is based on spider charts. A spider chart is composed of several radial spokes, one representing each acceptance indicator, which are structured by the DART dimensions (figure 1).

In order to evaluate the user acceptance, the developed indicators should be quantified and normalized on a scale from one to six. The value of one describes a high user acceptance and the value of six a low acceptance level respectively. The acceptance curve (bold black line in the figure) represents the average acceptance level of the interviewed persons (statistical median). The statistical spread resulting from the spread of the survey could be used to draw a surface (utilizing the lower and the upper quartile) (Amberg et al., 2004).

The used presentation is analogous to the popular dart game, where a dart hitting the centre of the disc denotes the highest possible score (the highest possible acceptance level respectively). By means of this visualization approach, potential acceptance challenges and resistances could easily be identified, focused and eventually even removed.

RESEARCH METHOD

To adopt DART, it is necessary to identify the indicators, which influence the user acceptance of web-based aptitude tests. The proposed DART dimensions are the basis for the examination. The intention is to find aggregated indicators, which show a high significance. Due to a review of current literature (e. g., Amberg, Wehrmann, Zimmer, 2004; DIN33430, 2002; Kollmann, 1998; Reichwald, 1980; Schröder and Meszlery, 2003; Schönecker, 1982; Stiftung Finanztest, 2004) and a workshop with users of web-based aptitude tests, we identified the following aggregated acceptance indicators (table 2):

(Perceived) Costs	(Perceived) Usefulness	(Perceived) Ease of Use	(Perceived) Network Effects
• Amount of time to locate the test • Amount of time to complete the test • Intrusion into privacy • Risk of refusal with regard to a study or a job • Test fee • Obligation to pass the test • Costs for driving to the test location • Amount of time to drive to the test location • Costs for the internet connection • Transparency of the test costs	• Detailedness of the test result • Potential benefit for the test user • Advantage of a web-based aptitude test in comparison to the written or oral form • Cost-effectiveness of the test execution • Scale of evaluated competencies • Information content with regard to the evaluated competencies • Scale of statements about the test person • Sufficient information about the test before test execution • Satisfaction of information needs • Correctness of the test result	• Understandability of the test questions • Loading time of the aptitude test • Understandability of the test results with regard to the test questions • Pleasing text • Degree of visualization • Warranty of the authentic answers to the test questions • Menu navigation • Clarity of the test • Fault tolerance • Temporal flexibility during the test execution • Interactivity with regard to the analysis of the acquired data	• Communication of the test features in advance • Reputation of test supplier • Scale of background information • Up-to-dateness of the test result • Up-to-dateness of test implementation • Improved self-reflection • Admission to a study programme • Confidence in the test result • Improved choice of the study programme • Fun factor of the test execution • Storage of the test result

Table 2: Acceptance indicators of web-based aptitude tests

For the survey, we used a questionnaire consisting of 76 questions for the acceptance indicators. With regard to the DART method, we chose a scale from 1 (absolute approval) to 6 (absolute denial) as measures for the acceptance indicators. For evaluation, we select an external web-based aptitude test in the area of business. It supports applicants and students to choose their further working field. Therefore, we choose students with an appropriate subject (e. g., business administration, business information technology) for the sample. They were asked to try the test and fill out the questionnaire at the university or at home within four weeks in October 2005. We ended up with 75 total responses.

RESULTS

In order to verify the developed evaluation instrument, we firstly purified the measurement, then proved the reliability and afterwards conducted a factor analysis according to Churchill (1979 p. 66 ff.).

We analyzed the **item-to-total correlation** within every indicator. Items were eliminated if their corrected-item to total correlation (correlation of each item with the sum of the other items in its category) was less than .30. The key assumption for this procedure is that all items, if they belong to the domain of the concept, should have an equal amount of common core. If all the items in a measure are drawn from the domain of a single construct, responses to those items should be highly intercorrelated. Basically, a high item-to total-correlation is recommended. Scores of item-to-total correlation between .30 and .50 are average, above .50 high (Weise 1975). Items with a low correlation can't measure the construct well. To receive appropriate scores of reliability, we tested **Coefficient Alpha** at the same time and resolved according to Nunnally (1967) that Coefficient Alpha should be better than .70 (Nunally 1978, S. 245).

The analysis of the item-to-total correlation and the coefficient alpha results in the reduction of 29 acceptance indicators. Table 3 reports the remaining acceptance indicators, which have a higher coefficient alpha than .7 and a higher item-to-total correlation than .3.

Dimension	Coefficient Alpha	No	Acceptance indicators	Item to total correlation
(Perceived) Usefulness	.930	U1	Cost-effectiveness of the test execution	.820
		U2	Sufficient information about the test before the test execution	.832
		U3	Satisfaction of information needs	.864
		U4	Scale of evaluated competencies	.701
		U5	Information content with regard to the evaluated competencies	.719
		U6	Potential benefit for the test user	.841
(Perceived) Costs	.851	C1	Costs for driving to the test location	.740
		C2	Amount of time to drive to the test location	.740
(Perceived) Ease of Use	.703	E1	Understandability of the test questions	.401
		E2	Pleasing text	.658
		E3	Degree of visualisation	.434
		E4	Clearness of the test	.538
(Perceived) Network Effects	.820	N1	Improved choice of the study program	.563
		N2	Scale of background information	.713
		N3	Improved self-reflection	.764
		N4	Up-to-dateness of the test result	.554

Table 3: Remaining acceptance indicators

In the last step, a **factor analysis** was conducted using principal component analysis as the means of extraction and varimax as the method of rotation. Therefore, we tested the data with the "measure of sampling adequacy (MSA)" (Kaiser and Rice, 1974). The MSA-value of our data is .878. This means, that the data are adequate for the factor analysis. The analysis resulted in three factors instead of the four proposed factors: (Perceived) Usefulness, (Perceived) Ease of Use, (Perceived) Network Effects and (Perceived) Costs (cf. table 4).

Acceptance Indicators	Component		
	1	2	3
U1 Cost-effectiveness of the test execution	**.850**	-.030	-.048
U2 Sufficient information about the test before the test execution	**.853**	.066	-.190
U3 Satisfaction of information needs	**.887**	-.176	-.152
U4 Scale of evaluated competencies	**.746**	-.307	-.056
U5 Information content with regard to the evaluated competencies	**.746**	.016	-.161
U6 Potential benefit for the test user	**.903**	-.148	-.082
C1 Costs for driving to the test location	.532	.078	**.753**
C2 Amount of time to drive to the test location	.442	-.197	**.809**
E1 Understandability of the test questions	.332	**.675**	-.060
E2 Pleasing text	.518	**.663**	-.012
E3 Degree of visualisation	**.628**	.157	.110
E4 Clearness of the test	.390	**.648**	.105
N1 Improved choice of the study program	**.703**	-.181	-.101
N2 Scale of background information	**.778**	-.218	-.030
N3 Improved self-reflection	**.845**	-.242	-.086
N4 Up-to-dateness of the test result	**.665**	.157	-.154

Table 4: Factor analysis (Rotated Component Matrix)

Factor loadings for all indicators, which represent the correlation between the indicators and the respective factors, are higher than .55 and are thus considered high (Nunnally, 1994). Together, the three observed factors account for 68.23 %. The scree test plot verifies the presence of the three distinct factors having eigenvalues greater than 1. On the basis of these results, it is necessary to adapt the DART-approach and the proposed dimensions for the evaluation of the user acceptance of web-based aptitude tests. When the dimensions (Perceived) Usefulness and (Perceived) Network Effects are combined, their composite reliability was higher than either of the two (cf. table 3). Alpha for the combined dimension was .943. This dimension considers all aspects of the usefulness of web-based aptitude tests. Furthermore, the results of the factor analysis show that the acceptance indicator "E3 Degree of visualisation" has a high factor loading on this combined dimension.

In the following, we conducted a univariate analysis to evaluate the user acceptance. We only considered acceptance indicators, which have an influence on the user acceptance of web-based tests (cf. table3). With regard to the DART method, we determined the statistical median, the lower and the upper quartile and used these values to visualize the user acceptance, cf. figure 2

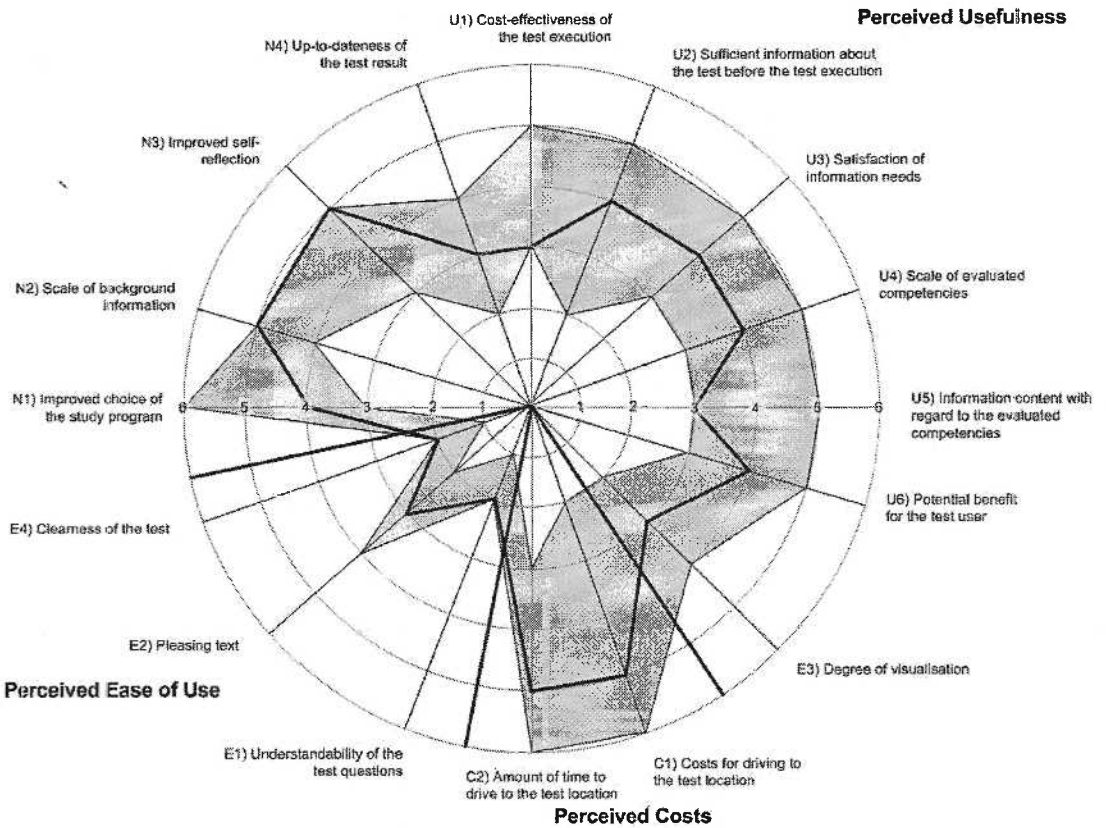

Figure 2: DART-Chart regarding the user acceptance of the evaluated web-based aptitude test

Figure 2 shows the heterogeneous importance of the three dimensions because each dimension is represented by a different amount of indicators. In this context, we adapted and reorganized the DART-chart. It clearly shows that the interviewees evaluated the evaluated indicators very differently. The interviewees would not invest time for the evaluated web-based aptitude test. This shows the acceptance indicator C2 of the dimension (Perceived) Costs. A reason could be that the (Perceived) Usefulness is not high enough. For example, the information content with regard to the evaluated competencies needs further improvement. Furthermore, detailed information is necessary for the choice of a study program. Nevertheless, the web-based aptitude test is user friendly designed. The (Perceived) Ease of Use is ranked well, the test questions are easy to understand and the text is appealingly worded.

SUMMARY AND OUTLOOK

The findings of this study suggest that there are three dimensions, which influence the user acceptance of web-based aptitude tests: (Perceived) Usefulness, (Perceived) Ease of Use, (Perceived) Network Effects and (Perceived) Costs. Furthermore, the results show that with the use of the DART-approach a detailed analysis of the influencing factors is generally possible. It helps to define a set of measurable acceptance indicators for the evaluation of the user acceptance.

The aim of our research is to develop an evaluation method for web-based aptitude tests. This method should focus the user acceptance of web-based aptitude tests and should help to understand the specific design criteria. The mistakes which can occur through the design of the web-based interface should be identified and minimized. The results of this study can be seen as a first step towards a systematically development of a web-based aptitude test. In addition, further questions of our research are: How can the result of such an evaluation be considered during the design process of a web-based aptitude test? How do personal factors influence the user acceptance and the evaluation process?

REFERENCES

Amberg, M., Hirschmeier, M. & Wehrmann, J. (2004). The Compass Acceptance Model for the analysis and evaluation of mobile services, *International Journal of Mobile Communications (IJMC)*, 2, 3, 248-259.

Amberg, M., Wehrmann, J. & Zimmer, R. (2004). Towards a Theory Based Identification of Influencing Factors for the design of mobile Services, in *Proceedings of the Conference on Cybernetics and Information Technologies, Systems and Applications*, July 21 - 25, Orlando, Florida, USA.

Churchill, G. A. Jr. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, 16, 64 - 73.

Davis, F. D. (1986). A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results. In *MIT* Sloan School of Management. Cambridge, USA.

Degenhardt, W. (1986). *Akzeptanzforschung zu Bildschirmtext: Methoden und Ergebnisse.* München: Fischer.

DIN33430 (2002). *Anforderungen an Verfahren und deren Einsatz bei der berufsbezogenen Eignungsbeurteilung: DIN 33430.* Berlin: Deutsches Institut für Normung.

Galletta, D. F. & Malhotra, Y. (1999). Extending the Technology Acceptance Model to Account for Social Influence: Theoretical Bases and Empirical Validation. In *Proceedings of the 32nd Hawaii International Conference on System Sciences*, Hawaii, USA.

Goodhue, D. L., & Thompson, R.L. (1995). Task-Technology Fit and Individual Performance, *MIS Quarterly*, 19:2 (6/1995), 213-236.

Hornke, L., & Zimmerhofer, A. (2005). Akzeptanz und Validierung eines webbasierten Studienberatungsinstrumentes. from http://www.studieneignung.de/.

Kaiser, H. F. & Rice, J. (1974). Little Jiffy, Mark IV. In *Educational and Psychological Measurement*, 34, 111-117.

Kaplan, R. S., & Norton, D. P. (1996). *The balanced scorecard – translating strategy into action.* Boston, USA: Harvard Business School.

Kollmann, T. (1998). *Akzeptanz innovativer Nutzungsgüter und -systeme: Konsequenzen für die Einführung von Telekommunikations- und Multimediasystemen.* Wiesbaden: Gabler-Press.

Nunnally, J. C. (1974). *Psychometric theory.* New York: McGraw-Hill.

Reichwald, R. (1980). Zur Notwendigkeit der Akzeptanzforschung bei der Entwicklung neuer Systeme der Bürotechnik. In Studiengemeinschaft AKZENTE (Eds.), *Akzeptanz neuer Bürotechnologie-Bedingungen für eine sinnvolle Gestaltung von Arbeitsplatz, Organisationsstruktur und Mitarbeiterbeteiligung.* Munich: University of the German Federal Armed Forces.

Schönecker, H. (1985). Forschungsprojekt Bürokommunikation im Auftrag des Bundesministers für Forschung und Technologie. In Picot, A.; Reichwald, R. (Eds.), *Bürokommunikation: Leitsätze für den Anwender*, Munich: CW- Publicationen.

Schröder, D., & Meszlery, K. (2003). Nutzerorientierte Kompetenzbilanzierung. Akzeptanz von Kompetenzmessverfahren. In *Unternehmen der Zukunft*, 4, Aachen, 18-20.

Stiftung Finanztest (2004). Eignungstest bestanden. In *Finanztest*, Zeitschrift der STIFTUNG WARENTEST, 11/2004, Berlin, 35-37.

Weise, G. (1975). *Psychologische Leistungstests.* Göttingen: Hogrefe.