



German Council for Social
and Economic Data (RatSWD)

www.ratswd.de

RatSWD

Working Paper Series

Working Paper

No. 126

Measuring Cognitive Competencies

Ulrich Trautwein

August 2009

Working Paper Series of the Council for Social and Economic Data (RatSWD)

The *RatSWD Working Papers* series was launched at the end of 2007. Since 2009, the series has been publishing exclusively conceptual and historical works dealing with the organization of the German statistical infrastructure and research infrastructure in the social, behavioral, and economic sciences. Papers that have appeared in the series deal primarily with the organization of Germany's official statistical system, government agency research, and academic research infrastructure, as well as directly with the work of the RatSWD. Papers addressing the aforementioned topics in other countries as well as supranational aspects are particularly welcome.

RatSWD Working Papers are non-exclusive, which means that there is nothing to prevent you from publishing your work in another venue as well: all papers can and should also appear in professionally, institutionally, and locally specialized journals. The *RatSWD Working Papers* are not available in bookstores but can be ordered online through the RatSWD.

In order to make the series more accessible to readers not fluent in German, the English section of the *RatSWD Working Papers* website presents only those papers published in English, while the German section lists the complete contents of all issues in the series in chronological order.

Starting in 2009, some of the empirical research papers that originally appeared in the *RatSWD Working Papers* series will be published in the series *RatSWD Research Notes*.

The views expressed in the *RatSWD Working Papers* are exclusively the opinions of their authors and not those of the RatSWD.

The RatSWD Working Paper Series is edited by:

Chair of the RatSWD (2007/ 2008 Heike Solga; 2009 Gert G. Wagner)

Managing Director of the RatSWD (Denis Huschka)

Measuring Cognitive Competencies

Ulrich Trautwein

University of Tuebingen (ulrich.trautwein[at]uni-tuebingen.de)

Abstract

The systematic of key cognitive competencies is of high scientific and societal relevance, as is the availability of high-quality data on cognitive competencies. In order to make well-informed decisions, politicians and educational authorities need high-quality data about the effectiveness of formal and non-formal educational environments. Similarly, researchers need strong data to test complex theoretical models about how individual biographies are shaped by the interplay between individual and institutional affordances and constraints.

Innumerable data sets offer some form of information on competencies such as respondents' years at school and their school grades. Such data are relatively easy to collect. When it comes to making informed political and educational decisions, however, there are increasing calls for a more systematic use of standardized competence tests. The production, storage, and use of standardized test data on competencies in specific domains is expensive, complex, and time-consuming, however.

This chapter argues that there is a paucity of adequate data on cognitive competencies in important domains, especially of longitudinal data from standardized competence tests, and that for many important questions there are no good alternatives to high-quality standardized tests of cognitive competencies. Furthermore, it outlines some challenges in the construction and application of standardized competence tests and makes several recommendations.

Keywords: cognitive competencies, assessment, intelligence, school grades

1. The Need for Systematic Assessments of Cognitive Competencies

Competencies are the ability to master complex demands in particular contexts. The many components of competent performance include knowledge, cognitive and practical skills, attitudes, emotions, values, and motivations (see Rychen and Salganik 2001; Weinert 2001, for a detailed definition of competencies). The scope of this chapter is restricted to cognitive competencies that are taught and learned in formal and non-formal learning environments. These cognitive competencies include, for instance, subject-specific knowledge, reading and mathematical literacy, computer literacy, and job-related knowledge.

The systematic and rigorous assessment of key cognitive competencies is of high scientific and societal relevance, as is the availability of high-quality data on cognitive competencies. It is now widely accepted that in modern knowledge societies the economic prosperity of individuals, communities, and countries is associated with the cognitive competencies acquired. In order to make well-informed decisions, politicians and educational authorities need high-quality data about the effectiveness of formal and non-formal educational environments. Similarly, researchers need strong data to test complex theoretical models about how individual biographies are shaped by the interplay between individual and institutional affordances and constraints. Some questions that require high-quality data on cognitive competencies include: Has high school students' mathematics and reading literacy generally increased or decreased in recent years and decades (see Becker et al. 2006)? Are *Abitur* standards and the related competence levels comparable across the German federal states (Trautwein et al. 2007)? Do female and male students and students from different family and ethnic backgrounds have the same access to high-quality education? Which domain-specific competencies are important for success in different domains at university and in the workplace (e.g., Nagy 2006)?

The need to assess and document the competence levels achieved by learners in formal and non-formal learning environments is generally accepted; innumerable data sets offer some form of information on competencies. For instance, the official statistics report the number of students who leave school with certain school-leaving certificates; school authorities document the distribution of school grades assigned in different grade levels and school types in each school year; and scientific studies ask students about their academic standing relative to their peers.

Such data on school-leaving certificates, grades attained in various learning environments, and self-reports of achievement are relatively easy to collect. They inform many questions and add to the body of knowledge about educational systems. When it comes to making informed political and educational decisions, however, there are increasing calls for a more systematic use of standardized competence tests. In the wake of international benchmarking studies such as PISA (Baumert et al. 2001), there has been increased scientific interest in Germany in the conceptualization, psychometric modeling, operationalization, and description of cognitive competencies. The production, storage, and use of standardized test data on competencies in specific domains is expensive, complex, and time-consuming, however. The question therefore arises of whether standardized competence tests might be substituted by cheaper and more readily available alternatives.

This chapter argues that there is a paucity of adequate data on cognitive competencies in important domains, especially of longitudinal data from standardized competence tests, and that for many important questions there are no good alternatives to high-quality standardized tests of cognitive competencies. The chapter is structured as follows. The next section provides a short description of standardized tests of cognitive competencies. Section 3 highlights the differences between domain-specific cognitive competencies and intelligence tests. Three frequently used but qualitatively different approaches to measuring cognitive competencies (grades or certificates; self-reports of competence; self-concepts) are compared and contrasted in section 4, and their advantages and disadvantages discussed. Challenges in the construction and application of standardized competence tests are subsequently outlined. Finally, several recommendations are made. For the sake of brevity and based on data availability, the chapter draws primarily on data collected in schools and universities.

2. Standardized Tests of Cognitive Competencies

Standardized tests of cognitive competencies use students' responses to certain stimuli (or "items") to infer competence levels. Carefully constructed standardized assessments such as those used in the PISA study are based on a conceptual model of what is being assessed, and their construction and evaluation is informed by psychometric models and state-of-the-art statistical analyses. Psychometrically constructed standardized tests have to fulfill a number of criteria. Most importantly, they must be objective (i.e., the resulting test scores must be independent of the person who administers and scores the test), reliable (i.e., the test must be internally consistent and give consistent results over time), and valid (i.e., the test must

actually measure what it sets out to measure). When standardized tests with high validity are used, the competence levels of all test takers can be compared directly, independent of where they live or their learning environment. Some well-known standardized tests of cognitive competencies include the TOEFL test assessing the English-language skills of non-native speakers and the PISA tests assessing verbal, mathematical, and scientific literacy.

Tests can be distinguished along several dimensions. *Curriculum-oriented tests* are based on material defined in the learners' curriculum. For instance, a curriculum-oriented mathematics test would implement tasks covered in the mathematics curriculum. In contrast, tests such as those implemented in PISA, which are based on the *literacy concept*, probe for competencies considered essential for full participation in society. Ideally, tests of cognitive competencies allow comparison across test takers ("*norm-referenced tests*") and inform on the individual test taker's absolute level of competence ("*criterion-referenced tests*" or tests with "*competence levels*").

Standardized tests such as PISA have helped to close the knowledge gap surrounding the cognitive competencies of various student groups (e.g., boys vs. girls; students with different immigration status). For instance, findings have shown that immigrant groups are differentially successful in different school systems. When carefully constructed, standardized tests have a high degree of fairness because all students receive a similar "treatment". Moreover, it is possible to discern items that may place some subgroups at a disadvantage and to eliminate these items from the test.

A specific advantage of carefully constructed standardized competence tests is that they allow the development of competence to be tracked over time. Forms of "anchoring" allow tests scores to be compared longitudinally, provided that the conceptual model is good and the quality of measurement is high.

To date, longitudinal data on the development of cognitive competencies over time are in short supply in Germany. Although some data sets contain such information, they tend to be relatively small, restricted to some areas of Germany, and/or the competence tests used are of limited quality (Blossfeld 2008). The National Educational Panel Study (NEPS; Blossfeld 2008) commencing in 2009 will help to overcome this deficit by examining students' mathematics, verbal, scientific, and ICT literacy as well as their literacy in English as a foreign language. Furthermore, some domain-specific tests will be administered to subgroups (e.g., business students).

3. Cognitive Competencies vs. Intelligence

The construction of psychometrically sound tests of domain-specific cognitive competencies is complex and expensive. Some critics have questioned whether these efforts are strictly necessary or whether cheaper alternatives are available. One proposed alternative is to use measures of general, decontextualized cognitive dispositions, such as intelligence (Rindermann 2006). Rindermann claimed that the competence tests used in large-scale assessment studies such as TIMSS and PISA measure a single cognitive ability that is practically identical to general intelligence. Given the relatively high intercorrelations observed between mathematics literacy, reading literacy, and cognitive ability, it might therefore be argued that it would be easier and cheaper to use intelligence tests instead of tests of domain-specific competencies in large-scale assessments.

This line of argumentation has major limitations, however (Baumert et al. 2007). First, there are clear conceptual differences between domain-specific cognitive competencies and general, decontextualized cognitive dispositions such as intelligence (e.g., in processes of knowledge acquisition and information processing and in dependence on the quality of educational environments). Second, although there is a statistically significant correlation between intelligence and scores on domain-specific competence tests, the results of construct validation studies provide strong empirical support for the multidimensionality (i.e., empirical separability) of cognitive measures applied in large-scale educational assessments (see Baumert et al. 2007). Third, evaluations of the educational effectiveness of a specific school, state, or country differ across domains, as shown, for instance, by a recent study (Trautwein et al. 2007) comparing educational outcomes at the end of the academic track in two German states (Baden-Württemberg and Hamburg). Although the Baden-Württemberg students clearly outperformed the Hamburg students in mathematics, with an effect size of Cohen's $d = .98$, the respective differences in English achievement ($d = .16$) and reasoning ($d = .07$) were negligible. Fourth, intelligence and domain-specific competencies differentially predict academic outcomes such as success at university (Nagy 2006).

Taken together, domain-specific cognitive competencies are theoretically and empirically separable from general, decontextualized cognitive dispositions such as intelligence, which are less amenable to educational interventions (see the expertise by Stern, in this volume). Tests of intelligence cannot replace psychometric tests of cognitive competencies in assessments of educational effectiveness.

4. Other Measures of Cognitive Competencies

4.1 Grades and (School-Leaving) Certificates

Many data sets contain information on teacher-assigned school grades and/or (school-leaving) certificates. For example, the official statistics in Germany document in detail a broad range of certificates acquired in formal education (e.g., school-leaving certificates; university diplomas; completed apprenticeships). Similarly, many data sets contain information on teacher-assigned grades or teacher evaluations of student progress (e.g., school grades; university grades). Without question, grades and certificates affect individuals' academic biographies and long-term success on the job market, and thus represent important information that should be documented. However, to what extent can these easily available data replace information obtained using complex and expensive standardized achievement tests? Three aspects are critical here: reference group effects, the association between background variables and teacher-assigned grades, and the reliability of self-reports.

4.1.1 Reference Group Effects: Restricted Comparability

Prior research has clearly documented that achievement scores collected via standardized achievement tests correlate only moderately with teacher-assigned school grades (Baumert et al. 2003; Ingenkamp 1971). Although teacher-assigned grades typically give a rather accurate (but not perfect) estimate of the position of each student within a class, teachers' differential grading standards mean that grades do not typically provide a valid basis for gauging achievement across classes or schools. The majority of teachers in Germany and in many other education systems do not use an absolute criterion for achievement when assigning grades (as is the case in standardized achievement tests). Rather, they tend to grade on a norm-referenced basis (Ingenkamp 1971), with the best student in the class receiving a very good grade and the weakest student a bad grade or a "fail." As a consequence, "grading-on-a-curve" effects can be observed in most schools in Germany. The size of the correlation between school grades and standardized competence tests typically ranges from about $r = .30$ to $r = .60$ (e.g., Baumert et al. 2003; Trautwein et al. 2007). When individual achievement is controlled, higher class-average achievement is associated with lower grades (Trautwein et al. 2006). Clearly, it is important to distinguish theoretically and empirically between these two indicators of achievement. Furthermore, teacher-assigned grades cannot easily be used to measure learning gains over time.

Given that school grades are not on a common metric across teachers and schools, it is hardly surprising that students who acquire the same school-leaving certificate in different schools or states do not necessarily exhibit the same level of cognitive competencies.

Moreover, although qualifications such as the *Hauptschulabschluss* (lower school leaving certificate) and *Abitur* (certificate giving access to higher education) are awarded at various school types in Germany, little is known about the comparability of these certificates across school types.

4.1.2 Effects of Sex, Family Background, and Immigration Status

Teacher-assigned grades have been shown to be influenced not only by cognitive competencies, but also by various student characteristics. Importantly, it is well documented that teachers' evaluations of students' cognitive competencies are associated with students' family background (e.g., Baumert et al. 2001) and influenced by teachers' gender stereotypes. Teachers' evaluations of students from immigrant families are also likely to be affected by stereotypes, but these effects may be compensated by grading leniency. More empirical studies are needed in this context.

4.1.3 Validity of Self-Report Data

Information on school grades and certificates can be collected via either self-reports or school records. Are students' self-reports of their grades reliable indicators of their actual grades or should school records be consulted? Several recent studies have reported high associations between self-reported and teacher-reported grades. For instance, Dickhäuser and Plenter (2005) reported a correlation of $r = .88$ for the last mathematics grade. It must be noted, however, that participants in these studies did not have anything to gain from reporting higher school grades than they actually attained. In a different context, the association between self-reported and teacher-reported grades may well be lower.

4.2 *Self-Assessments of Cognitive Competencies: Restricted Validity and Group Differences*

A quick, easy, and direct approach to assessing cognitive competencies is to ask individuals for an "objective" evaluation of their own competencies. For instance, students might be asked to report their competencies in logical reasoning or grammar (e.g., Kruger & Dunning 1999) and these self-ratings then correlated with data from a standardized test or an expert rating. As shown in a meta-analysis by Mabe and West (1982), the resulting associations are typically moderate in magnitude and vary from study to study. Mabe and West were able to identify some characteristics of studies that moderate the association between self-ratings and other indicators of competencies. Higher associations are found, for instance, if respondents expect their self-reports to be compared with objective evaluations and if some guarantee of anonymity is given in the study instructions. Even under such favorable conditions, however,

the associations between self-reported competencies and external information on these competencies were far from perfect. Looking at various domains, moreover, Kruger and Dunning (1999) showed that people with low abilities in these domains were particularly likely to overestimate their abilities. Kruger and Dunning attributed these misjudgments partly to the lower metacognitive competencies of these respondents. Taken together, the validity of self-evaluations of cognitive competencies is restricted.

4.3 Domain-Specific Self-Concepts

Domain-specific self-concepts are another frequently used construct in many empirical studies. Domain-specific academic self-concepts reflect a person's self-evaluation regarding a specific academic domain or ability (see Trautwein et al. 2006). These self-concepts are usually collected via self-report measures. Typical self-concept items are "I am quite good at mathematics" (mathematics self-concept) and "I have a poor vocabulary" (verbal self-concept). Although self-concepts share some similarities with self-evaluations of competencies, there is one crucial difference. Self-concept instruments ask specifically for a person's subjective self-evaluation, not for an "objective" self-evaluation. It is therefore not surprising that these instruments elicit external frame of reference effects (e.g., respondents compare their accomplishments with those of their friends or schoolmates rather than using an "average" comparison group) as well as internal frame of reference effects (e.g., respondents compare their competencies in mathematics with their competencies in English), yielding a complex pattern of associations with other assessments of competencies. Domain-specific self-concepts have proved to be predictive with regard to the competency development. However, they are no substitute for standardized tests of cognitive competencies.

4.4 The Need for Multiple Indicators: A Research Example

Which indicator of cognitive competence is the best predictor of a successful transition from school to university or the labor market? Modern educational systems work on the assumption that competence levels predict future success in higher education and the workforce. However, it has also been argued (e.g., Solga 2005) that employers rely heavily on the type of school-leaving certificate as a "signal" when hiring apprentices or employees. These certificates are more easily accessible than, for instance, test scores, and may thus have more influence in determining applicants' professional success than their actual level of competence. There is indeed some reason to believe that – given their easy availability to employers – school-leaving certificates and school grades have more pronounced effects on

success in the application process, whereas competencies predict success during vocational training and occupational careers. As plausible as this reasoning may seem, however, there is a need for empirical studies that empirically tease apart the confounding effects of certificates and competence levels cross-sectionally and longitudinally. Moreover, there is a need to distinguish among facets of cognitive abilities. Some studies from the United States seem to indicate that general ability (intelligence) plays a more important role in training success than do specific competencies (e.g., Ree and Earles 1991). Convincing empirical support for such a pattern of results is still lacking in the German context, however, primarily because of the lack of data sets including information on competencies measured by standardized tests as well as teacher-assigned grades and certificates.

5. Standardized Competence Tests: Challenges

The empirical assessment of competencies is more difficult than it may appear at first glance. Theoretically and empirically sound competence models are required as a basis for the development of measurement procedures. The systematic integration of theoretical frameworks, psychometric models, and measurement approaches often necessitates interdisciplinary cooperation, which introduces another level of complexity. The challenges facing longitudinal competence measurement outlined below are among those currently being addressed.

As yet, there is disagreement on which *domains of cognitive competencies* can be meaningfully measured by standardized tests and how differentiated the measurement should be. These questions are, for instance, very relevant to job-related cognitive competencies. Similarly, as regards to *criterion-referenced tests and competence levels*, there is also some disagreement on which competence levels can be considered sufficient, which levels can the majority of learners realistically achieve, and who should be responsible for establishing these standards in different domains.

Another challenge pertains to *possible positive and negative effects of competence testing*. What are the effects of systematic competence assessment in learning environments? For instance, do teachers make changes to the learning content covered or to their methods of teaching in response to the introduction of competence tests, and are the overall effects positive or negative? Moreover, in standardized educational assessments such as PISA to date, *unmotivated test taking* might have been the exception rather than the rule in Germany

(Baumert and Demmrich 2001). However, it is not clear if this may change in the future if standardized competence tests are administered more frequently.

Finally, *longitudinal measurement* is one of the most difficult and crucial challenges in the context of competence testing. Challenges include choosing an appropriate linking procedure, possible retest effects, the danger of ceiling and floor effects, and the question of whether the construct being measured remains the “same” over time (e.g., are multiplication tables in elementary school and complex numbers at upper secondary level part of the “same” mathematics?).

6. Recommendations

- a) Broader use of standardized tests of cognitive competencies can help to evaluate the effectiveness of educational institutions. Wherever feasible, standardized competence tests – in addition to or instead of measures such as teacher-assigned grades and self-assessed competence – should be used.
- b) Some effort should be put into investigating domains of cognitive competencies for which competence tests can be easily constructed (based on either the curriculum or the literacy concept) – and domains in which standardized competence tests are not feasible.
- c) There is a need for more high-quality tests that are available for researchers for use in their own projects (e.g., intervention studies). In this sense, there should be broader access not only to data but also to measurement instruments.
- d) When reporting “competence” data, researchers should always specify whether standardized tests or alternatives were used. It is especially important to critically address possible reference group effects and whether specific groups (e.g., gender or immigrant groups) might be at a disadvantage.
- e) Concerted efforts should be made to strengthen expertise in constructing and interpreting standardized competence tests in the scientific and non-scientific communities. There has been considerable progress in recent years (e.g., the German Research Foundation’s (DFG) priority program on “Competence models for recording individual learning outcomes and for reviewing educational progress”), but more expertise is needed across a broader population of researchers.

- f) Ways must be found of linking competence data collected in empirical studies (e.g., school achievement studies) with other data sets (e.g., data available from national agencies).

References:

- Baumert, J./Brunner, M./Lüdtke, O. and Trautwein, U. (2007): Was messen internationale Schulleistungsstudien? – Resultate kumulativer Wissenserwerbsprozesse. Eine Antwort auf Heiner Rindermann [What do international student achievements measure? – Results of cumulative learning. A reply to Heiner Rindermann.] *Psychologische Rundschau*, (58), 118-129.
- Baumert, J. and Demmrich, A. (2001): Test motivation in the assessment of student skills: The effects of incentives on motivation and performance. *European Journal of Psychology of Education*, (16), 441-462.
- Baumert, J./Klieme, E./Neubrand, M./Schiefele, U./Schneider, W./Stanat, P./Tillmann, K.-J. and Weiß, M. (Eds.) (2001): PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich [PISA 2000: Student literacy in international comparison]. Opladen, Germany.
- Baumert, J./Trautwein, U. and Artelt, C. (2003): Schulumwelten: Institutionelle Bedingungen des Lehrens und Lernens [School environments: Institutional conditions of learning and instruction]. In: Baumert, J./Artelt, C./Klieme, E./Neubrand, M./Prenzel, M./Schiefele, U./Schneider, W./Tillmann, K.J. and Weiß, M. (Eds.): PISA 2000: Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland. Opladen, 259-330.
- Becker, M./Trautwein, U./Lüdtke, O./Cortina, K.S. and Baumert, J. (2006): Bildungsexpansion und kognitive Mobilisierung [Educational expansion and cognitive mobilization]. In Hadjar, A. and Becker, R. (Eds.): Die Bildungsexpansion. Erwartete und unerwartete Folgen. Wiesbaden, 63-89.
- Blossfeld, H.-P. (2008): Education as a lifelong process. A proposal for a National Educational Panel Study (NEPS) in Germany. Bamberg.
- Dickhäuser, O. and Plenter, I. (2005): "Letztes Halbjahr stand ich zwei": Zur Akkuratheit selbst berichteter Noten [On the accuracy of self-reported school marks]. *Zeitschrift für Pädagogische Psychologie*, (19), 219-224.
- Ingenkamp, K. (1971): Die Fragwürdigkeit der Zensurengebung [The dubiousness of school grades]. Weinheim.
- Kruger, J. and Dunning, D. (1999): Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, (77), 1121-1134.
- Mabe, P.A. and West, S.G. (1982): Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*, (67), 280-296.
- Nagy, G. (2006): Berufliche Interessen, kognitive und fachgebundene Kompetenzen: Ihre Bedeutung für die Studienfachwahl und die Bewährung im Studium [Vocational interests, cognitive and scholastic abilities: Their role in choice of major and success at university]. Unpublished doctoral dissertation, Free University, Berlin, Germany.
- Ree, M.J. and Earles, J.A. (1991): Predicting training success: Not much more than g. *Personnel Psychology*, (44), 321-332.
- Rindermann, H. (2006): Was messen internationale Schulleistungsstudien? Schulleistungen, Schülerfähigkeiten, kognitive Fähigkeiten, Wissen oder allgemeine Intelligenz? [What do international student assessment studies measure? School performance, student abilities, cognitive abilities, knowledge, or general intelligence?]. *Psychologische Rundschau*, (57), 69-86.
- Rychen, D.S. and Salganik, L.H. (2001): Defining and selecting key competencies. Göttingen.
- Solga, H. (2005): Ohne Abschluss in die Bildungsgesellschaft: Die Erwerbschancen gering qualifizierter Personen aus soziologischer und ökonomischer Perspektive [Entering the education society without qualifications. The employment opportunities of lower qualified individuals from the sociological and economic perspectives]. Opladen.
- Trautwein, U./Köller, O./Lehmann, R. and Lüdtke, O. (Eds.) (2007): Schulleistungen von Abiturienten: Regionale, schulformbezogene und soziale Disparitäten [School achievement of students at Gymnasium schools: Regional, track-specific, and social disparities]. Münster.
- Trautwein, U./Lüdtke, O./Marsh, H.W./Köller, O. and Baumert, J. (2006): Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth grade mathematics. *Journal of Educational Psychology*, (98), 788-806.
- Weinert, F.E. (2001): Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit [Comparative measurement of achievement in schools]. In F. E. Weinert (Ed.), *Leistungsmessungen in Schulen*. Weinheim, 17-31.