Psychological Test and Assessment Modeling, Volume 61, 2019 (2), 149-160

Item Response Theory (IRT) analysis of the SKT Short Cognitive Performance Test

Mark Stemmler[1], *Ferdinand Keller*[2] & *Raphaela Fasan*[3]

Abstract

In this article, the SKT Short Cognitive Performance Test (SKT = Syndrom Kurztest; Erzigkeit, 2001) was tested with models of IRT. Here, the SKT test functions as an example on how IRT-analyses with any cognitive test, which results in a summary score, can be applied. The SKT was newly scaled in 2015 (Stemmler, Lehfeld & Horn, 2015) using a regression-based continuous norming approach (Crawford & Garthwaite, 2006) while leaving the testing material unchanged. The data were from a sample of 811 older adults ($\overline{x}$ = 73.3 years; *SD*=7.2) assessed at multiple testing sites in Germany. The sample consisted of cognitively healthy older adults (*n* =250), of older adults suffering from mild cognitive impairment (MCI; *n*=290) as well as adults suffering from dementia (*n*=271). Analyses to check the psychometric validity of the SKT encompass the Partial Credit Model (PCM), the Graded Response Model (GRM) and Multidimensional Response Models (MRM).

The strict assumptions of the one parametric Rasch model with regard to specific objectivity were not met. The two parametric GRM provided a better goodness-of-fit. A two-dimensional GRM revealed that the subtests assessing *Memory* and those assessing *Attention* can be distinguished as two separate dimensions. However, the use of a total score seems still valid, but, next to a careful inspection of the total score the scores for *Memory* and *Attention* should be taken into account and considered separately.

Keywords: Syndrom Kurztest (SKT), short cognitive performance test SKT, Item Response Theory (IRT), Partial Credit Model (PCM), Graded Response Model (GRM), Multidimensional Response Models (MRM)

[1] *Correspondence concerning this article should be addressed to:* Prof. Dr. Mark Stemmler, Institut für Psychologie, Universität Erlangen-Nürnberg, Nägelsbachstrasse 49c, 91052 Erlangen, Germany; email: mark.stemmler@fau.de

[2] Klinik für Kinder- und Jugendpsychiatrie/Psychotherapie, Universität Ulm, Germany

[3] Institut für Psychologie, Universität Erlangen-Nürnberg, Germany

If a patient is suspected to suffer from a dementia due to Alzheimer's disease or any other dementia type, short cognitive tests like the Mini Mental State Examination (MMSE; Folstein, Folstein und McHugh, 1975), the Montréal Cognitive Assessment Test (MoCA; Nasreddine et al., 2015) or the short cognitive performance test SKT (= Syndrom-Kurztest; Erzigkeit, 2001) are usually employed to assess the global degree of cognitive impairment in order to support the diagnosis of a possible dementia (Deuschel & Maier, 2016; Stemmler & Kornhuber, 2018). The SKT was developed based on the methods of the classical test theory (CTT). The construction of a summary score follows a simple addition of the scaled scores; like for any other cognitive test the summary score indicates the degree of cognitive deterioration. The SKT's summary score ranges between 0 and 18 points; values above 10 are indicative of a possible pathological cognitive deterioration due to dementia.

The SKT (Erzigkeit, 2001) is a well-established and widely employed instrument for the assessment of cognitive impairment in older adults. The subtests assess two different cognitive functions. *Memory* is assessed through recall (immediate and delayed) and recognition of pictures of painted commonly used objects (subtests II, VIII and IX). The measured raw scores are the not memorized or not recognized objects. *Attention* is measured through speed of information processing and the respective time in seconds needed to complete each subtest (subtests I, III, IV, V, VI and VII). For example, the patient needs to arrange magnetic blocks on a board in ascending order or the patient should rearrange the blocks back to their original positions. The SKT's new scaling from 2015 utilized a continuous norming approach as suggested by Crawford and Garthwaite (2006). For the norming the raw scores are measured in either 'the number of not memorized objects' and the 'time needed to complete an attention-task'. The raw scores or the actual cognitive performance of a subject are compared to a target performance which is based on a regression equation; important predictors of the used multiple regressions were gender, age and intelligence. Intelligence was assessed based on either the total Wechsler Adult Intelligence Scale (WAIS-IV; German version from Petermann, 2012) or based on the mean of two subtests of the WAIS-IV. One subtest measured fluid intelligence (e.g., the matrices test) and one subtest assessed crystallized intelligence (e.g., the vocabulary test). If one divides the difference between the actual and the target performance by a *standard error of a new case* according to Crawford and Howell (1998), the resulting statistics follow a *t*-distribution with $df = n–2$. Negative differences, that is bad cognitive performance, below a certain percentile of the norming population were assigned deviation points. With the new norming we determined that percentiles below 25 and 16 are being significant deviations from the target performance. Positive differences, that is a better cognitive performance than predicted results a in deviation point of '0'. Negative differences, that is a worse cognitive performance than predicted will be assigned the value '1' if the resulting t-statistic is below the 25^{th} percentile. The value '2' will be assigned, if it is below the 16^{th} percentile. Finally, the new scaled scores of the nine subtests are summed up. The scaled scores of the regression based norming cannot be looked up in any tables; an analysis-program in EXCEL which is part of every delivered manual is needed. One needs to enter age, gender and intelligence and the raw scores for each subtest into the EXCEL sheet. Subsequently, the scaled scores including

Table 1:
Structure and overview of the SKT with regard to subtests and their measured cognitive functions

SKT Subtests	Labels of the Subtests	Cognitive Function
Subtest I	Naming Objects	Attention
Subtest II	Immediate Recall	Memory
Subtest III	Naming Numerals	Attention
Subtest IV	Arranging Blocks	Attention
Subtest V	Replacing Blocks	Attention
Subtest VI	Counting Symbols	Attention
Subtest VII	Reversal Naming	Attention
Subtest VIII	Delayed Recall	Memory
Subtest IX	Recognition Memory	Memory

Note. For the IRT analyses, the scaled scores of the subtests were treated as response categories of the items.

the summary score are presented by the evaluation program. The total summary score ranges between 0 ('healthy') and 18 ('pathological impairment'). An overview of the structure and the different subtests of the SKT can be taken from Table 1.

For the IRT-analyses below, the assigned normed scores of '0', '1' and '2' were treated as the three possible response categories of an item, here the 'responses' regarding the nine subtests.

Subtest IX *Recognition Memory* was excluded from the psychometric analyses due to an extremely skewed distribution of the multiple regression residuals which could not be smoothed properly through any transformation. If the residuals of the multiple regression for each subtest deviated substantially from the normal distribution, a transformation of the residuals were administered. The applied transformation for subtest I was 1 : x, for subtest III and V we applied a logarithmic transformation and for the subtests IV, VI and VII a square root transformation. The assessment of the two cognitive functions *Memory* and *Attention* was validated through factor analytic analyses (Overall & Schaltenbrand, 1992). But like any other cognitive test used for the assessment of cognitive impairment which measures several cognitive functions (e.g., the MMSE measures in addition to memory, orientation, attention and visuo-spatial skills), the final evaluation is always based on the summary score (Stemmler, Lehfeld & Horn, 2015). This implies that the subtests measuring *Memory* and *Attention* are based on one common underlying cognitive function or dimension. The goal of the following study was to take the SKT as a representative example on how to apply IRT-models on cognitive tests which result in a summary score to evaluate a subject's cognitive impairment. IRT-models are able to provide evidence of a one-dimensionality which is a prerequisite for the summation of scaled scores. In the following, Rasch models (Rasch, 1966; Kubinger, 1989) were tested using polytomous data (PCM; cf. Lang, 2016). In addition to the Rasch models also the

graded-response model (GRM) and Multidimensional Response Models (MRM) were investigated. This article can be seen as an hands-on example on how cognitive tests which use a summary score can be tested and possibly be revised based on the results of an IRT analysis.

Methods

This study was based on a sample of $n = 811$ patients, which were tested during 2005 and 2009 in several testing centers in Germany (for more details, the reader may consult Stemmler, Lehfeld, Siebert & Horn, 2017; Hessler, Stemmler & Bickel, 2016). The assessment of dementia was part of the primary health care routines of the hospitals involved using the ICD 10 criteria for a diagnosis (Dilling, Mombour & Schmidt, 2005). The sample consisted of 55.2 % women with age varying between 60 and 91 years ($M = 73.3$, $SD = 7.2$). Of the total sample, there were $n = 271$ patients diagnosed with a dementia of the Alzheimer‘s type and/or a vascular dementia. Patients suffering from another type of dementia or from a dementia of an unknown origin were excluded from the study. Another $n = 290$ patients were diagnosed as suffering from a mild cognitive impairment (MCI); $n = 250$ patients with no cognitive impairment, they were classified as being cognitively healthy.

In our analyses, we treated the three possible values of the scaled scores for each subtest as response categories in IRT-terms of an item. Therefore, the Partial Credit Model (PCM; Masters, 1982) for the eight subtests was applied. We assumed the summary scores from 0 to 18 as representing a one-dimensional construct of ‚cognitive impairment‘ with higher scores representing a more severe cognitive impairment. For each summary score a person parameter could be assigned; negative person parameters stood for a low cognitive impairment and positive person parameters for a severe cognitive impairment. The tripartite response set required two threshold parameters with an increasing order of the thresholds (threshold 1 differentiates the category (or deviation point) 0 from 1 and the next threshold with a larger value differentiates the category points 1 and 2). The mean of the threshold parameters (i.e., item location; cf. Rost, 1996; Alexandrowicz, Friedrich, Jahn & Soulier, 2015) reveals where on this one-dimensional continuum called ‘cognitive impairment’ each subtest is located. For testing the assumption of specific objectivity, the Likelihood-Ratio-Test (LRT, Andresen, 1973) was applied for low vs. high scorers for the total sample and for the demented patients, MCI and the healthy subjects separately. For testing the assumption of one-dimensionality of items, the Martin-Löf-Test (MLT; Christensen et al., 2002) was used. Here, two groups of items which supposedly measure different constructs are compared to each other. Therefore, the subtests are grouped according to the results of the factor analysis into a *Memory* (subtests II and VIII) and an *Attention* dimension (subtests I, III, IV, V and VI, VII).

Since it is common that the Rasch model with its strict assumptions (e.g., specific objectivity, equal item discrimination) only rarely fits empirical data, two-parametric IRT models were also applied. We used the graded-response model (GRM) for testing the

assumption of different item discrimination parameters. Furthermore, the GRM was used to test a solution with two different dimensions against the one-dimensional GRM.

For the analyses the R packages (R Core Team, 2009) were employed. For the estimation of the model parameters and for the model control of the PCM, the R package *eRm* (Mair et al., 2012) was applied. The two-parametric IRT-models were investigated with the R package *mirt* (Chalmers, 2012).

Results

Analyses with the Rasch model

Figure 1 reveals the person-item-diagram for the one-dimensional modelling of the total sample using the PCM. The person parameters happened to be in the value range between –2 and +1, therefore in the range of no to mediumly severe cognitive impairment. The lower part of the diagram displays horizontal bars for each subtest on which the thresholds are marked. The item locations are the bold dots, the thresholds are marked by circles. The ordering of the subtests top-down reflects the ranking according to the item difficulty, with the most difficult subtests being listed at the top. It is obvious, that all

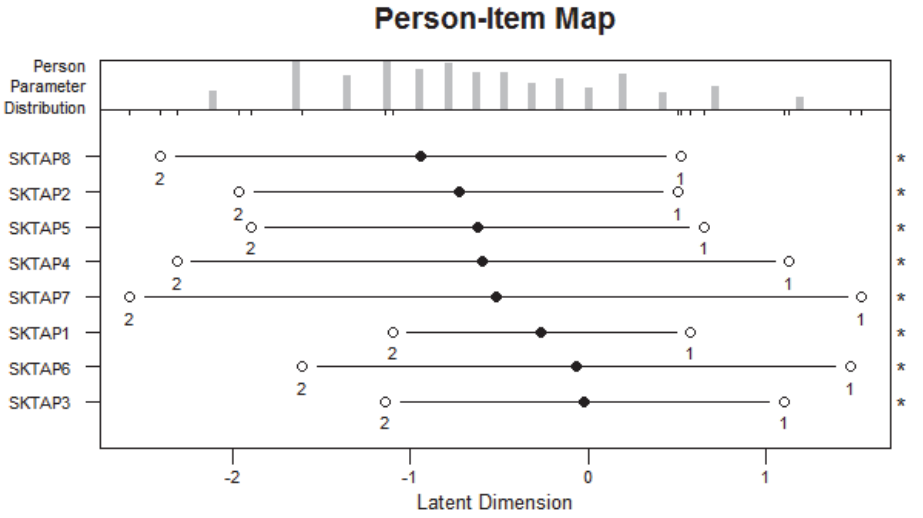

Figure 1:
Person-subtest-diagram for the total sample (n=811)

Note. Negative values of the latent dimension represent no cognitive impairment, positive values represent cognitive impairment. Subtests of the SKT: SKTAP1 = Naming Objects; SKTAP2 = Immediate Recall; SKTAP3 = Naming Numerals; SKTAP4 = Arranging Blocks; SKTAP5 = Replacing Blocks; SKTAP6 = Counting Symbols; SKTAP7 = Reversal Naming, SKTAP8 = Delayed Recall.

item locations are close-by and within a value range of –0.94 to –0.02 ($\bar{x}$ = –0.47, *SD* = 0.30). The two subtests measuring *Memory,* VIII *Delayed Recall* and II *Immediate Recall* were the most difficult subtests, while the subtests measuring *Attention,* VI *Counting Symbols* and III *Naming Numerals* were the easiest ones. The threshold parameters should be ordered monotonously in ascending order, such that threshold 1 marked the transition from the first response category to the second, and subsequently threshold 2 the transition from the second response category to the third. The asterisks at the margin of the figure display that the ordering of the thresholds is reversed in all items/subtests.

The significant result of the LRT for the total sample indicates that the item locations are different for groups with high and low scores (χ^2 = 107.81, *df* = 15, *p* < .05). Comparing the high with the low scoring subjects within the subgroups revealed that the LRT turned out to be significant for the MCI and demented subjects (demented χ^2 = 43.26, *df* = 15, *p* <. 05; MCI: χ^2 = 54.64, *df* = 15, *p* <. 05), but not for the healthy subjects (χ^2 = 17.42, *df* = 15, *p* = .29). Therefore, the requirements for specific objectivity were only partially met.

Concerning one-dimensionality of items, the MLT for the total sample turned out to be significant (χ^2 = 298.92, *df* = 47, *p* <. 05), indicating that there are characteristic differences between the *Memory* and *Attention* subtests. Also, the comparisons within each of the three subsamples turned out to be significant: healthy: χ^2 = 100.96, *df* = 47, *p* <. 05; MCI: χ^2 = 137.47, *df* = 47, *p* <. 05; demented: χ^2 = 116.04, *df* = 47, *p* <. 05.

Comparison of the subtest locations in the three subsamples

Based on Figure 2 one can see, that the item locations of the healthy subsample display a very specific pattern or profile, while the item locations of the MCI and demented subsample seem to be more similar. The item locations of the healthy subsample were above the item locations for the *Memory* subtests II *Immediate Recall* and VIII *Delayed Recall* and located below the item locations for the subtest VII *Reversal Naming* for the MCI or demented sample. The item locations of the subtests I, IV, V and VI were located within the confidence interval band; for the subtest III, the item location of the healthy sample was close to the confidence interval band of the demented sample. Valid subtests for all subsamples turned out to be the *Attention* tasks I *Naming Objects*, IV *Arranging Blocks*, V *Replacing Numerals* and VI *Counting Symbols*, here, the item locations for all subsamples were within the confidence interval band of the demented sample. It seems that healthy subjects perceive the item difficulty of some subtests differently than the MCI or demented subjects.

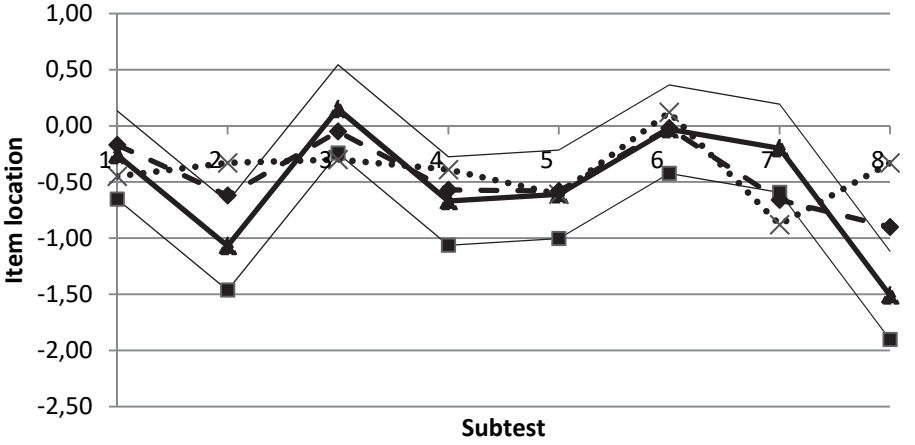

Figure 2:
Item locations of the eight subtests for each subsample (healthy subjects, persons with MCI and patients suffering from dementia)

Note. Bold solid line = demented; bold dashed line = MCI and bold dotted line = healthy subjects. ‚dementia ± 2 SE (thin line)' represents roughly a 95 %-confidence interval band; this was calculated by multiplying the standard error for each item location by 2. Subtests of the SKT: 1 = Naming Objects; 2 = Immediate Recall; 3 = Naming Numerals; 4 = Arranging Blocks; 5 = Replacing Blocks; 6 = Counting Symbols; 7 = Reversal Naming, 8 = Delayed Recall.

Analyses with the graded-response model

The PCM approach for testing the data of the total and the different subsamples revealed that specific objectivity was only partially met and especially the Martin-Löf-Test suggested two separate dimensions such as *Memory* and *Attention*. Thus, the PCM was compared to solutions revealed by two-parametric one- and two-dimensional IRT models. The decision for the best-fitting and most parsimonious model was based on the BIC (cf. Schwarz, 1978). Other information criteria such as the AIC, corrected AIC and sample-size adjusted AIC (cf. Chalmers, 2012) were in complete agreement with the respective solution favored by the BIC; thus, their values are not reported. Furthermore, the Likelihood-Ratio-Tests between models as provided by *mirt* were significant with $p < .0001$ except for GRM vs. GPCM (n.s.).

It appeared, that the GRM had a much better fit than the PCM when assessed by the BIC (GRM: *BIC* = 12627.9; PCM: *BIC* = 12780.5). The goodness-of-fit BIC of the Generalized Partial Credit Model (GPCM; Muraki, 1992) that was tested as an alternative to the GRM was very similar to the BIC of the GRM (GPCM: *BIC* = 12626.8). This supports

the assumption that the subtests have different abilities to differentiate between the subjects' cognitive performance.

A GRM with two dimensions even revealed a better fit (*BIC* = 12461.6). The discrimination parameters of the 2-dimensional GRM were transformed into factor loadings for a better comparability of the eight SKT subtests (see Table 2). The two *Memory* subtests (II and VIII) showed similar loadings on one factor, while the other subtests revealed high loadings on a second factor; subtest VII *Reversal Naming* did not load high on any of the two factors. The two factors are correlated with *r* =.32.

Table 2:
Factor loading from the two-dimensional graded response-model for the total sample

Subtest	Labels	Factor 1	Factor 2
SKT I	Naming Objects	.010	-.516
SKT II	Immediate Recall	-.975	.037
SKT III	Naming Numerals	.071	-.558
SKT IV	Arranging Blocks	-.034	-.802
SKT V	Replacing Blocks	.011	-.815
SKT VI	Counting Symbols	-.021	-.764
SKT VII	Reversal Naming	.118	-.342
SKT VIII	Delayed Recall	-.657	-.140

Note. Factor 1 represents *Memory*; Factor 2 represents *Attention*.

Conclusions

IRT models were applied for the first time to the SKT according to Erzigkeit (Lang, 2016; cf. Kubinger & Draxler, 2007). The newly-normed SKT assigns deviation points for each subtest depending on how much a person's cognitive performance deviates from the predicted or expected cognitive performance. Because for each subtest or item, three response categories can be assigned (0,1, and 2), a Partial Credit Model (PCM) for polytomous response sets was applied. First, the specific objectivity, that is, a homogenous 'answering of the SKT items' independent of person or item parameters was tested. The LRT tests between the low versus high scorers in the total sample and in the three subsamples (healthy, MCI and demented) turned out to be significant except for the healthy subsample. Therefore, the strict assumptions of the one parametric Rasch model with regard to specific objectivity were not met. The Martin-Löf-Test (MLT) turned also out to be significant which indicates characteristic differences between the *Memory* and the *Attention* tasks. Therefore, one-dimensionality of the items is rejected.

The threshold parameters were not always in ascending order. Non-ascending or reversed thresholds usually are indicative of problems with item formulation or of a blurred differentiation between the response categories, here the deviation points for

each subtest (Rost, 1996). Problems with item formulations can be excluded with regard to the SKT, but it can be assumed that the category 1 is not much frequented and therefore not as much differentiating between the cognitive abilities as expected. The latter was also elaborated on a theoretical basis by Wetzel and Carstensen (2014). The middle category or 1 deviation point represents a cognitive ability located in the lower part of the t-distribution, that is between the 25th percentile and the 16th percentile. The analyses showed that the ordering of the categories was not violated. It can be assumed that either an elimination of this category or a widening of this category would be more appropriate in the future.

Within the IRT paradigm the subtests of the SKT do not meet the strict assumptions of the Rasch model. The two-parametric approach of the Graded Response Model (GRM) displayed a much better goodness-of-fit, suggesting a varying differentiation between the subtests. Furthermore, the two-dimensional GRM confirms that the cognitive performance for *Memory* and *Attention* should be differentiated. In the output of the EXCEL sheet of the newly-normed SKT, the scores for *Memory* and *Attention* are presented in different graphics, next to the total summary scores. Therefore, the total summary score still can be used, but it should be used and interpreted together with the scores for *Memory* and *Attention*.

The results of this article show the necessity of evaluating common cognitive tests for the assessment of cognitive impairment with IRT models. IRT modelling may proof that many implicitly made assumptions during the test construction do not hold, especially with regard to the summation of scaled scores to a total score which is usually taken as an indicator of cognitive impairment used for the diagnosis of a dementia due to Alzheimer's disease or any other dementia type.

Acknowledgment

The authors express their gratitude to Rainer Alexandrowicz for his helpful comments of an earlier version of the manuscript. A part of the statistical analyses were based on the Master thesis of Raphaela Fasan (maiden name: Lang). The new regression-based norming of the SKT was supported by Dr. Willmar Schwabe Arzneimittel, Karlsruhe, Germany.

References

Alexandrowicz, R., Fritzsche, S., & Keller, F. (2014). Die Anwendbarkeit des BDI-II in klinischen und nicht-klinischen Populationen aus psychometrischer Sicht. Eine vergleichende Analyse mit dem Rasch-Modell [The applicability of the BDI-II in clinical and non-clinical populations from a psychometric point of view. A comparative analysis with the Rasch model]. *Neuropsychiatrie [Neuropsychiatry], 28*, 63-73.

Alexandrowicz, R. W., Friedrich, F., Jahn, R., & Soulier, N. (2015). Using Rasch-models to compare the 30-, 20-, and 12-items version of the general health questionnaire taking four recoding schemes into account. *Neuropsychiatrie, 29*, 197–191.

Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika, 38*, 123-140.

Chalmers, R. P. (2012). mirt: A Multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48,* 1–29.

Christensen, K. B., Bjorner, J. B., Kreiner, S., & Petersen, J. H. (2002). Testing unidimensionality in polytomous Rasch models. *Psychometrika, 67(4)*, 563-574.

Crawford, J. R. & Garthwaite, P. H. (2006). Comparing patients' predicted test scores from a regression equation with their obtained scores: A significance test and point estimate of abnormality with accompanying confidence limits. *Neuropsychology, 20*, 259-271. doi:10.1037/0894-4105.20.3.259.

Crawford, J. R., & Howell, D. C. (1998). Regression equations in clinical neuropsychology: An evaluation of statistical methods for comparing predicted and obtained scores. *Journal of Clinical and Experimental Neuropsychology, 20*, 755-762. doi:10.1076/jcen.20.5.755.1132.

von Davier, M. (1997). WINMIRA - program description and recent enhancements. *Methods of Psychological Research Online, 2(2)*, pp. 25-28.

Deuschl, G., & Maier, W. (2016). S3-Leitlinie Demenzen [S3 Guideline Dementia]. In D. G. f. Neurologie (Ed.), *Leitlinien für Diagnostik und Therapie in der Neurologie [Guidelines for Diagnostics and Therapy in Neurology]*.

Dilling, H., Mombour, W., & Schmidt, M. H. (2005). *Internationale Klassifikation psychischer Störungen* [International Classification of Mental and Behavioural Disorders] (Weltgesundheitsorganisation Hrsg.) [World Health Organization Editor]. Bern: Huber.

Erzigkeit, H. (2001). *SKT: Kurztest zur Erfassung von Gedächtnis- und Aufmerksamkeitsstörungen – Manual [Short test for the assessment of memory and attention deficits - Manual]*. Erlangen: Geromed.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-Mental State" a practical method for granding the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 189-198.

Hessler, J. B., Stemmler, M. & Bickel, H. (2016). Cross-Validation of the Newly-Normed SKT for the Detection of MCI and Dementia. *Journal of Gerontopsychology and Geriatric Psychiatry*. DOI 10.1024/1662-9647/a000154.

Kubinger, K. D. (1989). Aktueller Stand und kritische Würdigung der Probabilistischen Testtheorie [Current status and critical appraisal of probabilistic test theory]. In K. D. Kubinger (Hrsg.), *Moderne Testtheorie – Ein Abriß samt neuesten Beiträgen [Modern Test Theory - An Abstract with the Latest Contributions]* (2. Auflage [Second Edition], S. 19 – 83). Weinheim: Beltz.

Kubinger, K. D., & Draxler, C. (2007). Probleme bei der Testkonstruktion nach dem Rasch-Modell [Problems with the test construction according to the Rasch model]. *Diagnostica, 53*, 131-143.

Lang, R. (2016) *Die Überprüfung des SKT an klinischen und nicht-klinischen Populationen mit dem Rasch-Modell [An examination of the SKT in clinical and non-clinical populations with the Rasch model]*. Unveröffentlichte Masterarbeit Friedrich-Alexander-

Universität Erlangen-Nürnberg [Unpublished master's thesis Friedrich-Alexander-University Erlangen-Nuremberg].

Mair, P., Hatzinger, R., & Maier, M. J. (2009). Extended Rasch Modeling: The R Package eRm. PDF-Dateianhang zum Programmpaket eRm. 1-24. Zugriff am 08.04.16 unter: https://cran.r-project.org/web/packages/eRm/vignettes/ eRm.pdf

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47(2)*, 149-174.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*, 159-176.

Nasreddine, Z. S., Philips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of American Geriatrics Society, 53*(4), 695-699.

Overall, J. E., & Schaltenbrand, R. (1992). The SKT neuropsychological test battery. *Journal Of Geriatric Psychiatry And Neurology, 5(4)*, 220-227.

Pauli, L., Daseking, M., Petermann, F. & Stemmler, M. (2018). Zusammenhänge zwischen den kognitiven Leistungen in einem Demenzscreening (SKT) und in einem Intelligenztest (WAIS-IV) bei älteren Menschen: „Welche kognitiven Leistungseinbußen im Alter sprechen für einen möglichen pathologischen Abbauprozess?" [Relationships between cognitive performance in dementia screening (SKT) and an intelligence test (WAIS-IV) in older people: What cognitive performance losses in old age indicate a possible pathological deterioration process]. *Zeitschrift für Gerontologie und Geriatrie [Journal for Gerontology and Geriatrics], 51,* 266-274. DOI: 10.1007/s00391-017-1263-x

Petermann, F. (2012) (Hrsg.) Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV). Deutsche Bearbeitung [German Adaptation]. Frankfurt: Pearson Assessment.

R Development Core Team (2009). *Computer software manual*. Wien: R Foundation for Statistical Computing. Erhältlich unter http://www.R-project.org.

Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danmarks Pædagogiske Institut; 1960.

Rost, J. (1996). Lehrbuch Testtheorie Testkonstruktion [Textbook on Test Theory and Test Contruction]. Bern: Verlag Hans Huber.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461-464.

Stemmler, M., Lehfeld, H., & Horn, R. (2015). *SKT nach Erzigkeit [SKT According to Erzigkeit]. SKT Manual Edition 2015.* Spardorf: Geromed.

Stemmler, M., Lehfeld, H., Siebert, J. & Horn, R. (2017). Ein kurzer Leistungstest zur Erfassung von Störungen des Gedächtnisses und der Aufmerksamkeit - SKT Manual Edition 2015 – und der regressionsbasierte Ansatz [A short cognitive performance test for the assessment of memory and attention deficits - SKT Manual Edition 2015 - and the regression based approach]. *Diagnostica, 63(4)*, 243-255. DOI: 10.1026/0012-1924/a000178.

Stemmler, M. & Kornhuber, J. (2018). *Demenzdiagnostik. Kompendien Psychologische Diagnostik [Dementia Diagnostics. Compendia Psychological Diagnostics] – Band 16 [Volume 16]*. Göttingen: Hogrefe Verlag.

Wetzel, E. & Carstensen, C. H. (2014). Reversed thresholds in partial credit models: A reason for collapsing categories? *Assessment*, *21*(6), 765-774.

Wright, B.D. & Panchapakesan, N. A. (1969). Procedure for sample-free item analysis. *Educational and Psychological Measurement, 29*, 23–48.