Validity of the Severe Impairment Battery Short Version
Jos F De Jonghe, Roland B Wetzels, Ans Mulders, Sytse U Zuidema,
Raymond T C M Koopmans

To cite this version:
Validity of the Severe Impairment Battery Short Version

Jos F.M. de Jonghe PhD\textsuperscript{a}, Roland B. Wetzels MD\textsuperscript{b,d}, Ans Mulders MD\textsuperscript{c,d}, Sytse U. Zuidema MD PhD\textsuperscript{d} and Raymond T.C.M. Koopmans MD PhD\textsuperscript{d}.

\textsuperscript{a}: Geriatric Medicine, Medical Center Alkmaar, Alkmaar, the Netherlands

\textsuperscript{b}: Care center Rivierenland Foundation, Tiel, the Netherlands

\textsuperscript{c}: SVVE De Archipel, Eindhoven, the Netherlands

\textsuperscript{d}: Department of Nursing Home Medicine, Radboud University Nijmegen, Medical Center, the Netherlands

Running title: SIB-S
Corresponding Author: Jos F.M. de Jonghe, PhD, Geriatric Medicine, Medical Center Alkmaar 046, Wilhelminalaan 12, 1815 JD Alkmaar, The Netherlands. Tel: 0031-72-5483090; Fax: 0031-72-5482197, Email: j.de.jonghe@mca.nl

Key words:
Neuropsychological tests
Psychometrics
Dementia
Reproducibility of Results
Factor Analysis, Statistical

Word count: 3.141, Table count: 4, Reference count: 37, Abstract count: 241

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article (if accepted) to be published in the Journal of Neurology, Neurosurgery & Psychiatry editions and any other BMJPL products to exploit all subsidiary rights, as set out in our licence (http://jnnp.bmjournals.com/ifora/licence.pdf).
Abstract

Background. Efficient neuropsychological tests are needed to measure cognitive impairment in moderate to severe dementia.

Objective. To examine construct validity of the Severe Impairment Battery Short version (SIB-S) in nursing home patients with moderate to severe dementia, and to examine potential floor effects for the SIB-S.


Results. A total of 290 patients were included 264 of whom had complete SIB-S protocols. Internal consistency of the SIB-S was very high (Cronbach’s alpha=.97). Principle components analysis produced 3 factors, the first of which explained more than 50% of common score variance. Semantic memory items loaded highly on the first factor. Total SIB-S scores were associated with cognitive impairment (SIB-S - MMSE rho = .91, p<.001), and with functional dependency (SIB-S - ADL scale rho = -.61, P<.001). SIB-S total scores differentiated between dementia stages as measured with the Global Deterioration Scale (F=164.6 df:3,260, P<.001). Comparisons of SIB-S total score variance across patients with moderate to severe dementia and patients with below or above average Mini Mental State Exam scores, indicate absence of large floor effects.

Conclusion. In this first study examining an independently administered SIB-S, the scale proved to be a homogeneous and valid measure of cognitive impairment. The SIB-S is a short scale that can be used to assess moderately to severely demented patients, who may find it difficult to complete traditional, lengthier neuropsychological tests.
The Severe Impairment Battery (SIB) is a cognitive test developed specifically for assessment of patients who are cognitively too much impaired to be able to complete standard neuropsychological tests.\textsuperscript{1} The SIB has become an important outcome measure in clinical trials involving patients with severe dementia.\textsuperscript{2,6} It takes 30 minutes to complete the SIB and for some severely impaired patients, this time is at the upper end of their attention span. Recently, a short version of the SIB (SIB-S) was developed that takes 10-15 minutes to administer and it can be completed by more profoundly impaired patients.\textsuperscript{7} Though content validity of the SIB-S was established in the development cohort, validity has not been examined in an independent patient sample.

Several studies from different countries have shown that the 51 items SIB scale is a reliable and valid measure of cognitive functioning across patient groups with varying degrees of dementia severity.\textsuperscript{7-14} Results from three different studies show that SIB scale items are based on 4-8 underlying factors with a strong first factor explaining a large part of common score variance.\textsuperscript{7,11,15} Patients in the development cohort of the SIB-S were assessed with the 51 item SIB. In the final SIB-S version, 26/51 original SIB scale items were retained.\textsuperscript{7} SIB-S total scores were strongly associated with SIB and MMSE total-scores. A recent Korean study, conducting a secondary analysis of the 51 items SIB showed that the SIB-S was a reliable and valid measure for evaluating patients with severe dementia.\textsuperscript{16} However, none of these studies examined psychometric properties of an independently administered SIB-S. Ideally, a scale’s psychometric properties are retained in the short version derived from it in a way that both measure the same aspects of behavior or cognition. Factor structure invariance across patient samples is one way of examining whether two scales measure the same cognitive abilities, thereby establishing construct validity.
This study examined cognitive functioning as measured with the SIB-S in nursing home patients with moderate to severe dementia. To our knowledge, it is the first time construct validity was examined in an independent patient sample. We hypothesized that the SIB-S factor structure would be invariant to the original SIB factor structure. Acknowledging that every cognitive test may show floor effects along the continuum of dementia, our second hypothesis was that floor effects would be less evident for the SIB-S as compared to another well-known cognitive measure.

Methods

STUDY DESIGN

This was a cross-sectional, observational study. Cognitive impairment, dementia severity and functional dependence were assessed in nursing home patients, allowing for a comparison between measures. The study is part of the WAAL BEhavior in Dementia (WAALBED)- part II study, a longitudinal study on the course of neuropsychiatric symptoms in nursing home patients with dementia. Approval of the regional research ethics committee was obtained. Relatives or legal guardians of all patients gave fully informed written consent.

PARTICIPANTS

Nine nursing homes in the Netherlands participated in the study. Resident nursing home physicians identified potentially eligible patients by systematically screening all inpatients from 14 dementia special care units. Patients were considered for inclusion provided they had dementia (defined with international criteria and Dutch consensus guidelines\textsuperscript{17}) and were institutionalized for at least 4 weeks. Patients were ineligible if a life-threatening disease was
present. The second author (RW) independently checked eligibility against patients’ clinical notes.

MEASUREMENTS AND PROCEDURES

Consensus meetings were organized to instruct and to ensure that standardized procedures were applied for all measures and assessments. Resident nursing home physicians or psychologists assessed cognitive impairment using the SIB-S and Mini Mental State Examination (MMSE), and rated severity of dementia as defined by Global Deterioration Scale (GDS) criteria. The MMSE is a well-known screening test for cognitive impairment and scores range 0-30. The Global Deterioration Scale (GDS) consists of descriptions of seven major, clinically distinguishable dementia severity stages, ranging from ‘no cognitive decline’ to ‘very severe’. Licensed vocational nurses, who were assigned to individual patients and who were specifically instructed to observe patient behavior a few days prior to assessment, rated Activities of Daily Living (ADL) as defined by Inter Resident Assessment Instrument – Long Term Care Facilities section G (2005, version 07). The InterRAI LTCF-ADL is an observational scale for ADL-dependence, which measures resident self-involvement in the personal activities of daily life. It includes 4 of the ADL items of the InterRAI, each with 6 response categories and is scored based on a decision tree with 8 scale categories, ranging from 0 (independent) to 6 (totally dependent) and 8 (activity not seen). When first introduced, its reported psychometric properties were good to excellent (inter-rater reliability) and an internal consistency (Cronbach’s alpha) of .90. Validity and reliability of the ADL scale are established in a study including Dutch nursing home patients with dementia.
The SIB has been used as a cognitive outcome measure in many clinical trials involving patients with severe dementia. SIB-S item selection is based on factor analysis of original SIB scale items and subsequent consensus discussions between authors of the SIB. In the final version of the SIB-S 26/51 original SIB scale items were retained. All scale items except two are coded 0,1,2. The other two are coded 0,1 and the maximum possible score on the SIB-S is 50. The SIB is a highly reliable and valid cognitive test. Two authors (AM and JJ) performed modifications on a provisional SIB Dutch version to suit the aims of this study. Subsequently, the SIB-S Dutch version was translated back independently. The primary author of the original SIB-S approved the SIB-S Dutch version after adjusting a few minor details.

ASSESSMENT

A total of 76% of patients were assessed with the SIB-S and MMSE on the same day, 90.4% had both tests within the same week, and for the remaining patients cognitive tests were administered 8-47 days apart. All data were collected on standardized patient record forms.

STATISTICAL ANALYSIS

Statistical calculations were performed using SPSS for Windows, version 11 (SPSS, Inc. Chicago, IL).

Complete test protocols were those with no more that 10% missing values per cognitive test. Whenever there were few missing values modal test item scores were used as substitutes. The internal consistency reliability of the SIB-S was examined using Cronbach’s alpha. Calculation of the sample size was based on the assumption that 10:1 observations-variables
are required in order for the correlation matrix to be sufficiently stable to be used in factor analysis: in case of a 26 items test, 260 patients.

Extracted factors with eigen values $\geq 1.0$ were subsequently rotated orthogonally according to simple structure criterion and with the aim of clear factor interpretation. Item factor correlations determined assignment of the item to a factor, with a required minimum factor loading of .40, and provided the item-factor correlation was at least .10 higher than that with another factor. Construct validity was further evaluated by retaining SIB-S factor scores and using them as dependent variables in stepwise regression analysis with MMSE items as predictors. As some of the measures are ordinal scales and because scale scores were not normally distributed in this study, associations between SIB-S, MMSE and ADL total scales were examined using nonparametric Spearman rank correlations.

ONEWAY analysis of variance with Scheffe post hoc comparisons was used to examine group differences using cognitive test scores as dependent variables and dementia severity as independent variable. Diagnostic accuracy was examined by calculating the area under the curve (AUC) for SIB-S and MMSE total scores comparing GDS 4 and GDS 5, GDS 5 vs. 6 and GDS 6 vs. 7.

In the original SIB-S study, floor effects were examined by comparing mean SIB-S scores and variances in patients with MMSE scores below the 50th percentile MMSE total score as compared to those with scores above the 50th percentile. We repeated the analysis by calculating the t-test statistic for comparisons of mean scores and we used Levene’s homogeneity of variances test. Additionally, potential floor effects were examined by comparing mean SIB-S and MMSE scores and the homogeneity of score variances across
patient groups in GDS 4-7 (t-test, Levene’s test), and by examining proportions of patients who failed to take the test or who had a total test score of 0 using Chi² statistic.

Two-tailed $P$ values of <0.05 were considered to indicate statistical significance.

**Results**

A total of 290 eligible patients were included in the study. Twenty-six patients had more than 10% missing data because they refused or were just too ill to take the SIB-S, or because of other reasons. Of the remaining 264 patients 36 had a score of 0 on the SIB-S. A total of 20/290 patients had missing data for the MMSE and 75/270 had a score of 0 on the MMSE.

So, SIB-S total scores did not differentiate between 62/290 patients (21.4%) who were intended to be tested, as compared to 95/290 (32.8%) for the MMSE: a reduction of 34.7% (Chi-square 9.5, df:1, $P=0.02$).

Patients with missing values for the SIB-S total score did not differ from other patients on any of the demographic and clinical variables, except for vision impairment ($P=.004$), with more patients without vision impairment completing the SIB-S than those with mild to severe vision impairment.

As this is a study of the SIB-S and for reasons of simplicity, the data of 264 patients with complete SIB-S protocols are presented.
TABLE 1 about here

Demographic and clinical characteristics are presented in table 1. The study sample is typical of Dutch psychogeriatric nursing home patients, with a preponderance of older residents, many of whom female, with low educational attainment, and with advanced dementia.

TABLE 2 about here

CONSTRUCT VALIDITY

Average SIB-S total score was 25.9 (SD 17.3, range 0-50) and average MMSE total score was 7.1 (SD 6.6, range 0-25). Cronbach’s alpha coefficient for the SIB-S was very high (α = .97). SIB-S item-rest correlations all were higher than .55, except for item 12 (Remembering investigator’s name, item-rest correlation: .30).

Three principal components that explained 67.4% of score variance (table 2) were found. Scale items that reflect understanding gestures and verbal instructions, and items reflecting recognizing and naming objects loaded highly on the first factor, which was called ‘Aphasia-Agnosia’. The second factor consists of items reflecting writing and copy drawing and it was called ‘Apraxia’. The third factor consists of items reflecting memory for newly learned material and it was called ‘Episodic memory’. Notably, the first factor explains more than half of common variance, indicating the SIB-S is a homogeneous measure.

In regression analysis of SIB-S factor scores MMSE items reflecting verbal ability shared unique variance with the first SIB-S factor. MMSE items reflecting nonverbal performance...
were associated with the second factor and MMSE items reflecting episodic memory were associated with the third SIB-S factor.

TABLE 3 about here

CONCURRENT VALIDITY
Low scores on total SIB-S were associated with cognitive impairment as measured with the MMSE (Spearman rho = .91, p<.001) and with functional dependency as measured with total ADL scale (Spearman rho = -.61, P<.001).

SIB-S total scores were associated with dementia severity. Patients in GDS stage 7 had lower scores compared to patients GDS 6, and patients in GDS 6 had lower scores than those in GDS 4-5. Diagnostic accuracy of SIB-S as measured with the AUC was modest for mild to moderate stages of dementia, but it increased importantly for moderate to very severe stages (table 4).

TABLE 4 ABOUT HERE

FLOOR EFFECTS
SIB-S total score variances were unequal across patient groups in different stages of dementia severity (Levene’s test: F=22.9, df=3,260, P<.001) (table 4). Post hoc analysis showed that score variance was less in GDS 4 as compared to GDS 5 (F=10.8, df=2,312, P=.002), score variance was less in GDS 5 as compared to GDS 6 (F=32.1, df=6,738, P<.001) and variance was greater in GDS 6 as compared to GDS 7 (F=26.4, df=13,586, P<.001). Notably, in GDS 7 MMSE total score variance was about two thirds less than the score variance in GDS 4-6.
The 50th percentile MMSE total score was 6. The mean SIB-S total score for patients with a MMSE score ≤ 6 (n=129) was 11.2 (SD=11.8, range= 0-39). For patients with a MMSE score > 6 (n=135) the mean SIB-S was 40.0 (SD=6.8, range= 16-50). Additionally, the mean and variance of the SIB-S total score differed significantly between the two MMSE groups for each sample (two-sample test with unequal variances, t=24.2, df=202,750, P<0.001, Levene’s test for equality of variances, F=77.8, P<.001). These results indicate the SIB-S does not show strong floor effects.

Discussion

This study evaluated validity of the SIB-S in nursing home patients with mild to severe dementia. Though principle component analysis produced 3 factors, the SIB-S is in fact a unidimensional measure of cognitive impairment. Cognitive impairment as measured with the SIB-S was associated with dementia severity and with functional impairment. Moreover, it was associated with another cognitive screening test and the SIB-S showed only modest floor effects.

These findings may have implications for neuropsychological evaluations of patients with advanced dementia. Few patients failed to take this relatively short cognitive test or had a total score of 0. The proportion of patients that did score 0 or failed to take the test was almost 35% lower than that for the MMSE. The wide range of test scores found for patients in the severe stages of dementia opens up the possibility of testing those who are usually considered not testable, without risking high drop out rates due to lengthy assessment procedures.
Though most of our results are in accord with those of others, some appear not to be. Previous studies reported correlations between SIB-S, SIB and MMSE that range .68-.99. In the original SIB-S development study subjects, stratified by MMSE 0-4, 5-7, and >7 exhibited a range of scores on the SIB-S, which suggests that there were no important ceiling or floor effects on the SIB-S. That study selected an 8 factor model accounting for 59.1% of common variance in the original 51 SIB items (US patient sample) and a 4 factor model accounting for 54.2% of variance (French patient sample). A second order factor analysis of the 8 factor model was not performed. In a pilot study (n=48) of SIB subscale scores a 4 factor model was selected. A secondary analysis using SIB data from a clinical trial examining the effects of memantine, showed a very strong first factor (eigenvalue > 20). However, all of these studies assessed patients with the original SIB 51 items version and no patient was independently assessed with the SIB-S. In this study patients were assessed with the SIB-S and a 3 factor model of SIB-S items was selected. Score variance accounted for was 67.4%, which indicates good model fit. The first factor explained more than half of the score variance. Also, the first factor eigenvalue accounted for more than 3.5 times the common variance of the second factor eigenvalue, which all in all strongly suggests that the SIB-S is a unidimensional measure of cognition. Others, using similar dementia screening tests found 2-4 latent structures underlying scale items. It is of interest to note that few factors are found when subtest scores are used as indicators, e.g. all questions pertaining to orientation are bundled and result in a single subscale score, while more factors are found when all individual items receive scores, e.g. ‘which day is it’ or ‘which year’, etc. Using subscales or creating item bundles is in effect conducting a second order factor analysis. When interpreting our results in the light of previous studies, it should be acknowledged that differences in testing procedures such as the number of test items used in regular or short scale form or fatigue resulting from lengthy test sessions, may lead to diverging factor analysis results. Strengths of
this study are the sufficiently large number of included patients, rigorous assessment procedures, classification of dementia severity, and use of the actual SIB-S scale.

Semantic memory loss is typically present in the later stage of dementia and by and large this is what the SIB-S measures. There is ample evidence from neuropathological and imaging studies suggesting that medial temporal lobe (MTL) atrophy is an early sign of Alzheimer’s disease, that the lesions are associated with episodic memory deficits, and that the neuropathological changes spread out to other regions of the brain later. \(^{29-35}\) Though the MTL is activated by semantic verbal memory processing, \(^{36}\) semantic memory loss is eminent when regions outside the MTL become affected. \(^{37,38}\) Mildly impaired patients find it increasingly difficult to remember recent conversations or to recapture what they did the day before. When patients reach the later stages of dementia, not only do they have episodic memory problems, but general knowledge about the world is lost too. They may find it increasingly difficult naming familiar objects correctly, they may no longer be able to recognize objects or know how to use them. In the later stages of dementia patients may lose all ability for learning new material and episodic memory tests will show floor effects. Factor analysis of SIB-S scale items and correlational analysis with specific MMSE items suggests that most SIB-S scale items tap semantic memory processes and few items measure episodic memory. These findings underline content validity of a scale that was constructed to be used in moderate to severe dementia.

SIB-S scale construction is based on classic test theory principles. Future studies might want to use a modern approach, e.g., Item Response Theory (IRT). This type of analysis would be ideal to understand the value of each item of the SIB-S and how it contributed to understanding the individual’s impairment. Further, IRT can be used to improve a test. IRT is
based on the analysis of a continuum and is well suited for understanding how tests and their items perform on a compilation of items. Given that the factor analysis has already been done, IRT is a simple next step.

No gold standard exists for neurocognitive testing in clinical settings. Though many clinical trials including dementia patients use measures like the ADAS-cog or SIB, psychologists working with nursing home patients choose to use many different cognitive tests and neuropsychological batteries. Such lack of standardization hampers comparability of clinical data, e.g. when individual patients are transferred from one institution to another and it does not facilitate data pooling in nursing home research projects. Therefore, we recommend using the SIB-S as the standard, brief, neuropsychological examination in moderate to severe dementia.

Weaknesses of the study are the cross-sectional design. Longitudinal SIB data are available, but as yet none are for the SIB-S. Future studies may want to focus on the SIB-S as a measure sensitive to change, so that it can be used in clinical trials and in studies monitoring cognitive decline.

Efficient scales are needed to evaluate cognitive functioning and cognitive change beyond the mild stages of dementia. The short version of the SIB proved to be a valid and unidimensional scale associated with dementia severity and ADL dependency. Item content and brevity of the SIB-S strongly suggest that this is a test suitable for assessing severely impaired patients, who are no longer able to complete lengthier and more difficult neuropsychological tests.
Acknowledgements

We would like to thank Janssen-Cilag B.V. for kindly granting permission to use their provisional SIB Dutch version. We would also like to thank Tjerk Schoemaker, neuropsychologist, for his help translating back the SIB-S.

WAALBED II is funded by University of Nijmegen Department of Nursing Home Medicine, ZonMw, Foundation Johannes de Deo the Netherlands, Foundation for the Promotion of Science in Nursing Home Care (SWBV).
Reference List


Assessment Instrument for Activities of Daily Living, cognition and depression].


23. Tinklenberg J, Brooks JO, III, Tanke ED, Khalid K, Poulsen SL, Kraemer HC,
    Gallagher D, Thornton JE, Yesavage JA. Factor analysis and preliminary validation
    of the mini-mental state examination from a longitudinal perspective.
    *Int.Psychogeriatr.* 1990;**2**:123-34.

24. Talwalker S, Overall JE, Srirama MK, Gracon SI. Cardinal features of cognitive
dysfunction in Alzheimer's disease: a factor-analytic study of the Alzheimer's

    Mini-mental state examination and mental deterioration battery: analysis of the


27. Noale M, Limongi F, Minicuci N. Identification of factorial structure of MMSE
    based on elderly cognitive destiny: the Italian Longitudinal Study on Aging.

28. Jones RN, Gallo JJ. Dimensions of the Mini-Mental State Examination among


