ORIGINAL MANUSCRIPT

Sequentially Administered WAIS-IV Short Forms in Traumatic Brain Injury: Screening for Subnormal IQ and Intellectual Deterioration

Joseph J. Ryan[1], David S. Kreiner[1*], Samuel T. Gontkovsky[2,3] and Allison Myers-Fabian[4]

Ryan J J, Kreiner D S, Gontkovsky S T, Myers-Fabian A. Sequentially Administered WAIS-IV Short Forms in Traumatic Brain Injury: Screening for Subnormal IQ and Intellectual Deterioration. Int J Psychopathol Psych Diag. 2022;1(1):01-25.

Abstract

This investigation examines the utility of Wechsler Adult Intelligence Scale-Fourth Edition short form intelligence quotients (SFIQs) and abbreviated General Ability Indexes (GAIs) to discriminate normal from subnormal intellectual functioning and detect possible cognitive deterioration. Participants were 60 individuals with traumatic brain injury. Following the standard sequential order of subtest administration and depending on performance, assessment may be terminated after two, three, four, or five subtests. Classification accuracy statistics indicated that all short form predictions exceeded the cognitively subnormal base rate (i.e., IQ = 35%; GAI = 28%), with hit rates from .83 to .93. Two-subtest SFIQs and estimated GAIs were not recommended. A three-subtest short form prediction exceeded the base rate for intellectual deterioration (Base Rate = 45%, Hit Rate = .80). A four-subtest short form was recommended for estimating Full Scale Intelligence Quotients; three- and four-subtest short forms were good predictors of the GAI, and a three-subtest short form was useful for identifying intellectual deterioration.

Key Words: *WAIS-IV; Intelligence; Classification accuracy; Short form; TBI*

[1]*Department of Psychological Science, University of Central Missouri, Warrensburg, Missouri*

[2]*Department of Psychology, Twin Valley Behavioral Healthcare, Columbus, Ohio*

[3]*Department of Psychiatry and Behavioral Health, The Ohio State University, Wexner Medical Center, College of Medicine, Columbus, Ohio*

[4]*Department of Neuropsychology, PLLC, Austin, Texas*

**Corresponding author: David S. Kreiner, Professor, Department of Psychological Science, School of Nutrition, Kinesiology, and Psychological Science, University of Central Missouri, Lovinger 1111, Warrensburg, MO 64093, USA, Tel: (660) 543-8076, E-mail: kreiner@UCMO.edu*

Received: July 29, 2021, Accepted: October 17, 2021, Published: February 25, 2022

Introduction

Neuropsychological evaluations are frequently requested in both outpatient and inpatient healthcare settings, and the Wechsler intelligence scales have been a common component of such evaluations since [1] developed the "hold" and "don't hold" tests [2]. When interview and testing time are limited and a precise intelligence quotient (IQ; Table 1 provides a key to all abbreviations.) is not required, practitioners may rely on brief measures of adult intelligence, such as the Reynolds Intellectual Assessment Scales [3], Kaufman Brief Intelligence Test-Second Edition [4], or Wechsler Abbreviated Scale of Intelligence-Second Edition [5]. However, because many practitioners wish to avoid the expense of purchasing brief intelligence batteries and the fact that the Wechsler scales have been the gold standard for the measurement of adult intelligence for 80 years, many practitioners simply ignore the competition and elect to administer short forms (SFs) of the current Wechsler Adult Intelligence Scale-Fourth Edition [6]. Other factors that reinforce this preference are the unique characteristics of each assessment situation and the findings of empirical research. Using a Department of Veterans Affairs Medical Center patient sample, [7] compared the WASI [8] and WAIS-III [9] and concluded that "if the clinician's goal is to obtain an accurate estimation of general intellectual functioning, the current results suggest that the WASI should not be used in the assessment of individual patients" (p 22).

When administrating a WAIS-IV SF, the criteria for subtest selection usually are based on time requirements, patient physical capabilities, and the efficacy of a subtest to echo the abilities most relevant to the referral question. The decision of which subtests to include in any given SF is typically made prior to starting the assessment. However, an alternative approach is to administer two, three, four, or five subtests in the sequence prescribed by the publisher [10], terminating the examination at the point (*e.g.*, after three subtests) where the examinee's scores are inconsistent with the presence of intellectual deterioration or subnormal general ability. Following completion of an assessment, scores on the 10-subtest administration might be requested for disability determination or insurance reimbursement. In this case, the remaining subtests can be completed without concern for order of administration effects.

In the United States between 2007 and 2011, the average number of traumatic brain injuries (TBIs) treated in emergency departments was well above two million [11]. If past experience is a gauge of future events, the number of TBIs requiring medical attention will not decrease but will show steady, annual increases [11]. A significant portion of these individuals will exhibit cognitive deficits and be referred for neuropsychological evaluation, which ideally includes an assessment of intellectual functioning. Administration of an intelligence scale is considered a routine component of the neuropsychological assessment of patients with TBI [12], provides a statistically- and normatively-based tool for determining the presence of intellectual deterioration (Advanced Clinical Solutions) [13], and delivers important information for maximizing the interventional efficiency of the rehabilitation psychologist [14]. For adolescent and adult referrals, the most widely used intelligence scale and the one preferred by neuropsychological clinicians is the WAIS-IV [15]. The 10 core subtests of the WAIS-IV may require more than an hour

TABLE 1
Key for Abbreviations

Abbreviation	Meaning
IQ	Intelligence Quotient
SF	Short Form
SFIQ	Short Form IQ
GAI	General Ability Index
FSIQ	Full Scale IQ
WAIS–III	Wechsler Adult Intelligence Scale–III
WAIS–IV	Wechsler Adult Intelligence Scale–IV
WAIS	Wechsler Adult Intelligence Scale
WASI	Wechsler Abbreviated Scale of Intelligence
WASI–II	Wechsler Abbreviated Scale of Intelligence–II
TBI	Traumatic Brain Injury
WAIS–R	Wechsler Adult Intelligence Scale–Revised
MR	Matrix Reasoning
LNS	Letter-Number Sequencing
SS	Symbol Search
VP	Visual Puzzles
FW	Figure Weights
CA	Cancelation
DS	Digit Span
PA	Picture Arrangement
CO	Comprehension
IN	Information
VC	Vocabulary
SI	Similarities
BD	Block Design
CD	Coding
AR	Arithmetic
ROC	Receiver Operating Curve
PC	Picture Completion
HR	Hit Rate
GCS	Glasgow Coma Scale
LOC	Loss of Consciousness
PTA	Post Traumatic Amnesia
ANOVA	Analysis of Variance

to administer to healthy examinees [6] and considerably longer when utilized with patients referred clinically for neuropsychological assessment [16]. Moreover, in most inpatient settings testing time is limited by declining reimbursement schedules and competing appointments with other diagnostic and treatment services. In outpatient and private practice situations, most reimbursement systems also demand a cost- and time-efficient approach to neuropsychological assessment. To deal with these realities, practitioners identify ways to obtain an appropriately comprehensive examination that fits the externally imposed time limitations. If, as mentioned previously, it is likely that an IQ estimate will be the primary contribution of the intellectual evaluation, it makes sense to administer a WAIS-IV SF so that ample time remains for the clinical examination of other cognitive and adaptive functions.

The selected subtest approach to Wechsler scale SF development and clinical application should be informed by the latest empirical research. However, there is reason to believe that the vast SF literature published during the past 70 years [17-19] may not be fully pertinent to the WAIS-IV. The reasoning behind this assertion is that the original Wechsler-Bellevue Intelligence Scale Form I [20] has undergone four major revisions which produced the WAIS [21], WAIS-R [22], WAIS-III [9] and WAIS-IV [6]. These iterations have included revised item content, expanded norms, and modifications of administration and scoring procedures. The WAIS-III and WAIS-IV revisions also involved the updating of theoretical foundations via the increased measurement of fluid reasoning, working memory, and processing speed constructs, the designation of subtests as core, optional, or supplemental, and significant structural alterations of the scale on the basis of factor analytical research. Thus, the traditional Verbal and Performance IQs are a thing of the past, now replaced by the Verbal Comprehension, Perceptual Organization, Working Memory, and Processing Speed indexes. To enhance clinical utility, there are also new composite scores for the WAIS-III and WAIS-IV labeled the General Ability Index (GAI) and the Cognitive Proficiency Index. Also, the WAIS-III introduced the new Matrix Reasoning (MR), Letter-Number Sequencing (LNS), and Symbol Search (SS) subtests, and the WAIS-IV added three more subtests (*i.e*., Visual Puzzles [VP], Figure Weights [FW], and Cancellation [CA]), significantly modified another subtest (*i.e*., Digit Span [DS]) and deleted two of the original subtests (*i.e*., Picture Arrangement [PA] and Object Assembly). The WAIS-III Full Scale IQ includes the Picture Completion (PCm), PA, and Comprehension (CO) subtests, while the WAIS-IV IQ deleted these components and replaced them with VP and SS. The 1997 scale requires 11 subtests to compute the Full-Scale IQ (FSIQ), while the 2008 edition FSIQ is established on 10 core subtests. Based on these changes, one must acknowledge that past SF research with the WAIS, WAIS-R, and even the WAIS-III may not fully generalize to the WAIS-IV.

Another concern is the recent investigation that provided evidence against the generalizability of WAIS-III SF research to the WAIS-IV. [23] reported that the best factor-based four-subtest SFs using the Taiwan WAIS-IV standardization sample differed from those derived from the Taiwan WAIS-III norm group [24]. The recommended tetrads identified for the WAIS-III exhibited lower relative validity and were less accurate in predicting individual WAIS-IV

Full Scale IQs. These findings were attributed to the major structural and content modifications made to the WAIS-IV, prompting Chen and Hua to recommend that new SF investigations be conducted with the 2008 revision. Because the vast majority of SF administrations will be in clinical settings, this also implies a particular need for research with patient samples and, more importantly, those from specific diagnostic groups (*e.g.*, TBI, multiple sclerosis, Parkinson's disease, and the various dementia syndromes). The present investigation reported below represents a first step to remedy this situation because it focuses exclusively on patients with TBI.

For over a decade, the WAIS-IV has been the gold standard for intellectual assessment, but only a few publications are available to assist clinicians with the selection and interpretation of SFs. Moreover, it is noted that the WAIS–IV research reported in the present paper is the first to use a sequential administration procedure with a TBI sample. In the meantime, a potentially important contribution to SF research was made by [25], who developed "psychometric effectiveness" rankings for all 25 two-subtest combinations derived from the 10 core subtests. When considered along with the referral question and unique characteristics of the examinee, these rankings are intended to assist practitioners in selecting the most appropriate dyad for a specific purpose. The dyad rankings were developed using predominantly Caucasian (63.5%) and African American (32.8%) men. All members of this diagnostically mixed sample were referred for neuropsychological evaluation at an urban Veterans Affairs medical center. The effectiveness rankings utilize the reliability, validity, and inter-class correlation coefficients as well as the magnitude of the differences between the FSIQ and each estimated short form IQ (SFIQ). Also included in the effectiveness rankings are the proportions of estimated IQs for each dyad that fall within $\pm$ 10 points of the Full Scale. Girard et al. did not provide tables to convert sums of scaled scores into FSIQ and GAI estimates for selected dyads. Therefore, a follow up investigation by [26] was conducted using a cross-validation design and a predominantly Caucasian (83%), diagnostically mixed sample referred for neuropsychological evaluation. Regression-based equations were generated to predict estimated FSIQ and GAI values. Conversion tables were constructed for the nine possible dyads derived from the Information (IN), Vocabulary (VC), Similarities (SI), Block Design (BD), VP, and MR subtests. Correlations between the actual and predicted values for both composites were impressively high for each of the nine dyads. However, across all SFs the number of estimates that fell within $\pm$ 10 points of the actual composites was less remarkable. Approximately 20% of the SF estimates of FSIQ and 16% of the estimated GAIs fell outside the specified 10-point interval. Denny et al. used the $\pm$ 10-point interval to illustrate the amount of error around individual predicted scores but did not specify how this value was selected. Perhaps they were following the lead of the Girard et al. "psychometric effectiveness" study?

Chen and Hua [23] examined the psychometric and clinical utility of all possible factor-based tetrads using the Taiwan WAIS-IV standardization sample. From 13 of the 10 core and five supplemental subtests (CO and CA were omitted), 90 SFs were created, each containing one subtest from the individual factor-based indexes of Verbal Comprehension, Perceptual Reasoning, Working Memory, and

Processing Speed. Chen and Hua used a number of techniques to determine the best four-subtest combinations including administration time, reliability, corrected part-whole correlations, paired samples *t*-tests of the mean and variance differences between the estimated and actual FSIQs, and the number of SFIQs falling within the 95% confidence interval of the actual Full Scale. All SFIQs were derived using the linear equating procedure recommended by [27]. Five SFs were recommended for clinical use, with the combination of IN + VC + DS + Coding (CD) emerging as the best of the lot. This tetrad produced SFIQ estimates that fell within (a) the 95% confidence limits of the Full Scale in 69% of the cases and (b) the same [28] ability classification (*e.g*., borderline, low average, average, *etc*.) in 68% of the cases.

A sample of adults living in the Netherlands was administered the 10 core subtests of the WAIS-IV Dutch Edition [29]. These individuals were referred for psychometric assessment and carried neurological diagnoses such as epilepsy, neurofibromatosis, chromosomal abnormalities, and TBI. From each participant's protocol, six SFs were extracted and the sums of scaled scores converted to SFIQ estimates according to the [27] method. Subtest selection was based on psychometric findings from previous studies and yielded five SFs composed of four subtests and a single SF consisting of three subtests. Analyses were run on the total sample focusing on SF reliability, part-whole correlations, and intra-class correlations. Also examined were the proportions of SFIQs falling within the 90% confidence interval of the Full Scale and comparisons of the SF and FSIQ means. In a second analysis, the total sample was divided into subgroups according to general ability level (FSIQ ≤ 79 or ≥ 80) and all analyses were repeated. For the total sample and individuals with below average ability (FSIQ ≤ 79), the best SF was VP + SS + VC + Arithmetic (AR). Among persons with higher ability (FSIQ ≥ 80), substituting MR for VP produced the most accurate four subtest SFIQ.

Fan and associates [30] developed a four-subtest SF (BD + IN + CD + AR) using the WAIS-IV, Chinese version. Study participants included the mainland China standardization sample and a clinical group composed of patients diagnosed with schizophrenia, schizoaffective disorder, obsessive compulsive disorder, or mild to moderate intellectual disability. The subtests were selected from previously administered full WAIS-IV protocols based on administration time requirements, correlations with their respective index total scores, and clinical sensitivity. SF deviation quotients were developed according to the equipercentile equating method. For the standardization sample and the separate patient groups, the SFIQs were highly correlated with the FSIQs and the split-half, test-retest, and intra-class coefficients were described as excellent. In the combined sample of controls and patients, the SFIQ fell within two standard errors of measurement of the Full Scale in approximately 66% of the cases. Also of note was the ability of the SF to discriminate individuals in the combined normal and patient sample with FSIQs ≤ 89 from those with average or higher (IQ ≥ 90) global intelligence. The classification accuracy statistics were: Sensitivity = .96%; Specificity = .84; Positive Predictive Value = .71; and Negative Predictive Value = .98. Results of a receiver operating characteristic (ROC) analysis yielded an Area Under the Curve of .90.

Meyers and colleagues [31] conducted a cross-

validation study of the Ward seven-subtest SF [32] which is composed of BD + SI + DS + AR + IN + CD + PC. This SF was originally developed for the WAIS-R and later applied to the WAIS-III [33]. The samples under study were predominantly Caucasian (92.1%), diagnostically heterogeneous, and referred for clinical neuropsychological examination. Seventy patients completed the supplemental PC subtest and the 10 core subtests from which a linear regression equation was used to derive SFIQs. Thirty-two additional participants constituted a cross-validation sample. Regression-derived short form composites were compared to estimates based on proration. Both methods were highly accurate in predicting FSIQ, although proration was more accurate than regression in this regard. In a subsequent investigation [34], the 10 standard WAIS-IV subtests along with the supplemental PC subtest were administered to French speaking patients with schizophrenia or schizoaffective disorder. From the completed protocols, two versions of the seven-subtest SF were extracted. The first included the BD subtest as originally prescribed by Ward [32]; whereas, the second SF substituted MR for BD. SFIQ estimates were derived by proration and linear regression. Results indicated that proration yielded more accurate FSIQ estimates than did the regression method. This study gave strong support for SF estimation of global ability (*i.e.*, FSIQ) in appropriate situations but cautioned against the utilization of the prorated index scores for measuring different aspects of intelligence.

The utility of the Ward seven-subtest SF was also examined using the Scandinavian version of the WAIS-IV [35] in a non-clinical sample 18 to 74 years of age. Participants spoke fluent Swedish and were administered PC as well as the 10 core subtests (*M* FSIQ = 115.7; *SD* = 11.6). The investigation assessed (a) the power of the seven subtests recommended by [32] to estimate FSIQ, (b) whether or not SF validity differed according to age levels, and (c) the possibility of further reducing the number of subtests in the SF without decreasing predictive accuracy. In the total sample, regression analysis indicated that the SFIQ explained 93.1% of the variance in the Full Scale. Subsequent analysis of the differences between the means of the FSIQs and the prorated SFIQ estimates (M SF IQ = 115.6, SD = 11.1) was not significant. To accomplish the second goal, the sample was divided into young (< 55 years, *M* age = 31.79 years, *SD* = 9.55) and older (≥ 55 years, *M* age = 65.12 years, *SD* = 4.58) groups for separate analyses. Results for the young group revealed that the SFIQ accounted for 93.0% of the variance in FSIQ and that the mean for the prorated SFIQs did not differ from the mean of the Full-Scale scores. In the older group, the SF accounted for 92.2% of the variance in FSIQ but the prorated average composite significantly underestimated the Full-Scale mean. The authors suggested that the seven-subtest SF may not be the most appropriate abbreviation for use with individuals 55 years and older. The third goal of the investigation indicated that the accuracy of FSIQ predictions dropped off precipitously when fewer than four subtests constituted a SF. The best four subtest combination was BD + SI + AR + CD. The prorated SFIQ accounted for 86.1% of the variance in the Full Scale, while the former composite significantly overestimated the latter.

Persons with relapsing-remitting multiple sclerosis often experience a significant degree of fatigue during extended neuropsychological testing. To address this problem, eight- and four-

subtest SFs were developed to estimate WAIS-IV FSIQs and GAIs, respectively [36]. The eight-subtest SF was created by excluding VC and BD from the 10 core subtests. These tasks were selected because they are among the most time demanding, with average administration requirements of approximately 14 and 12 minutes, respectively [16]. The four-subtest SF consisted of IN + SI + MR + VP, which are the subtests that remain in the GAI after VC and BD are eliminated. Using data on 47 outpatients (Caucasian = 46, women = 40), significant differences between the standard FSIQ and GAI means and the prorated SF means were absent. Moreover, the percentage of prorated SFIQs that fell within ± 3 points of the FSIQs was an impressive 100%, with only six individuals (12.8%) demonstrating a difference of three points. The four-subtest SF was less precise in estimating the GAI, but still performed well with 100% of prorated SF composites falling within ± 5 points of the standard GAI.

If a SF is employed as part of a neuropsychological battery in order to obtain an estimated FSIQ and/or GAI, subtests are almost always selected prior to conducting the assessment. For instance, in order to quickly (*i.e.*, 20 to 25 minutes) obtain a FSIQ estimate while simultaneously assessing verbal concept formation and visual-spatial organization, the SI + BD dyad might be an appropriate selection. If the clinician wishes to provide coverage of the four underlying WAIS-IV factors, perhaps the tetrad of IN + DS + VP + CD would suffice. A limitation of this approach is that it requires the examiner to violate the recommended sequence of subtest administration, which is intended to reduce fatigue and maintain a high level of interest in the various test items. If individual subtests are administered out of sequence, the WAIS-IV Administration and Scoring Manual [6] recommends that the irregularity be taken into consideration when interpreting the results. Unfortunately, no assistance is given on how to accomplish this feat.

For examinees physically capable of taking the standard WAIS-IV, SFs can be selected by administering the subtests in sequential order. Depending upon the examinee's performance, testing may be terminated after completion of two, three, four, or five subtests. If subnormal intelligence is suggested, the remaining subtests can be completed immediately following administration of a SF. Thus, subtests are selected during the examination, not prior to meeting the examinee. [10] assessed the four possible sequentially administered SFs in a sample of outpatients (86% Caucasian) with mixed medical, psychiatric, or neurological disorders. Because (a) selected subtest SFs rarely yield exact IQ or GAI estimates and (b) the vast majority of SF publications include a cautionary statement that an abbreviated test is intended for screening purposes only and should not be used when a precise IQ is required, Ryan et al. evaluated the SFs without using the traditional statistical approach (*e.g.*, *t*-tests and correlations) or determining the number of cases falling within one or two standard errors of the FSIQ (or GAI). Instead, cutoff scores (*i.e.*, IQ ≤ 69 or ≤ 79) were used to identify, at two base rates for impairment (*i.e.*, 14% and 34%), the likelihood of subnormal versus normal intelligence. These base rates reflect the frequencies of occurrence of each cutoff score, respectively, in the total sample of outpatient cases. With this information in hand, classification accuracy statistics were calculated to reflect the clinical utility of each SF. Results indicated that the SFs performed differently depending on which

cutoff and base rate values were specified. With the lower IQ and base rate, Sensitivity of the four SFs was generally poor (*i.e*., .53 to .65) for intellectual impairment but Specificities, Positive Predictive Values, Negative Predictive Values, and Hit Rates were all excellent (*i.e*., .90 to .99). When the higher IQ and base rate statistics were used, Sensitivities of the three-, four-, and five-subtest SFs were strong to excellent (.88 to .90) and Specificities, Positive Predictive Values, Negative Predictive Values, and Hit Rates were all strong to excellent (.88 to .99). This investigation demonstrated that SFs derived from a sequentially administered WAIS-IV can be highly effective when testing time is limited and the identification of an examinee as either intellectually impaired or not impaired is sufficient. For practical purposes, the three-, four-, and five-subtest SFs were all satisfactory for screening the two levels of general ability, especially when using the cutoff ≤ 79 and an impairment base rate of 34%.

The foregoing review demonstrates that WAIS-IV selected subtest SFs, regardless of their impressive psychometric characteristics, do not provide precise estimates of the FSIQ or GAI. Therefore, more research is needed that relies on classification accuracy statistics to determine the tangible utility of SFs as screening techniques for intellectual impairment. It would also be worthwhile to determine if SFs can help detect the likelihood of cognitive deterioration in individual patients. A research-based method to establish a patient's probable level of preinjury cognitive functioning is essential in this regard as the effects of TBI depend on the premorbid level of functioning as well as the type and extent of injury [37]. Deterioration may be identified by contrasting postinjury test results with a preinjury estimate based on demographic information such as age, education, gender, ethnicity, and preinjury occupation [13]. Also, none of the investigations discussed above focused on patients with a documented history of TBI. Yet, the high prevalence of TBIs and the fact that the WAIS-IV is the most frequently used assessment tool in the United States [15] demonstrates an acute need for research that uses this instrument with this population. Finally, four of the SF studies reviewed above were conducted in the United States using mainly Caucasian participants representing a variety of psychiatric, medical, and neurological disorders [10] [25-26] [31]. A fifth study conducted in the United States focused exclusively on patients with multiple sclerosis, the vast majority of whom were Caucasian women [36]. The remaining investigations were from Europe [34-35] [29], Taiwan [23], and the People's Republic of China [30]. The latter samples were described as healthy men and women or individuals with mixed neurological disorders, low IQ, or various psychiatric syndromes. Obviously, there is a need for SF investigations using American samples of persons with TBI.

The present study was undertaken to compare FSIQs and GAIs derived from a standard 10 core subtest administration of the WAIS-IV with estimated IQs and estimated GAIs based on the separate combinations of two, three, four, and five subtests and two, three, and four subtests, respectively. The SFs were derived by the sequential administration approach recommended by [10], which totally or partially eliminates the possibility of confounding fatigue factors with order of administration effects when obtaining SFIQs and estimated GAIs. It also allows the examiner to decide as the examination unfolds whether or not a standard WAIS-IV administration is needed.

Behavioral observations, performance below expectations based on past achievements, and qualitative findings on specific subtests [38] should be sufficient for determining a need to continue testing. Thus, when the purpose of the SF administration is to rule out subnormal intelligence, average or above average scores on BD and SI in the absence of abnormal responses (*e.g.*, broken gestalts on BD and/or over-inclusive or concrete answers on SI) suggest that the WAIS-IV testing can be terminated after completion of a two-subtest short form.

In order to improve on past SF research, the sample under study was ethnically diverse and composed exclusively of individuals with a documented history of TBI. The goal was to assess the clinical utility of each SF for determining whether an individual was likely to have normal (*i.e.*, FSIQ ≥ 80) or subnormal (*i.e.*, FSIQ ≤ 79) general intelligence subsequent to TBI. The subnormal cutoff was selected because it represents the high end of the Borderline level of functioning and is at or below that of approximately 92% of the WAIS-IV standardization sample. A second goal was to determine if SFIQs were useful for identifying patients with probable intellectual deterioration. SFIQs were compared with pre-injury demographically-based FSIQ estimates obtained from the Advanced Clinical Solutions for the WAIS-IV and WMS-IV [13]. It is worth examining the possibility that sequentially administered WAIS-IV SFs are useful for the identification of both subnormal intellectual functioning and the presence of cognitive deterioration.

Method

Participants

Sixty individuals (men = 47, women = 13) from inpatient and outpatient services of major urban trauma centers located in the Southeastern (n = 47) and Midwestern (n = 13) United States volunteered for the study. All participants gave informed consent and volunteered for the investigation. Each was initially evaluated by emergency medicine specialists followed by referral to other physicians (*e.g.*, neurologist, neurosurgeon, and/or neuroradiologist), as needed. Inclusion criteria were: (a) documented history of traumatic brain injury with availability of an initial Glasgow Coma Scale score [39] and/or official records documenting loss of conscious (LOC) and/or posttraumatic amnesia (PTA), (b) clinically assessed to ensure that each participant demonstrated clear consciousness, intact orientation, and was no longer experiencing posttraumatic amnesia, (c) proficient and literate English speaker with a functional dominant hand, and (d) no history of psychiatric, neurological, or learning disorder. The sample consisted of 23 Euro Americans, 19 Hispanic Americans, 14 African Americans, and 4 other ethnicities. Means for age and education were 35.20 years (SD = 14.40, Range = 18 to 75) and 13.43 years (SD = 3.12, Range = 3 to 22), respectively. Forty individuals (67%) had abnormal CT scans (*e.g.*, contusion, hemorrhage, hypodensities, etc.), while the initial GCS injury classifications were 23 mild, 9 moderate, and 28 severe. Of the mildly injured cases, 14 had abnormalities on CT scan and may be designated as having mild, complicated injuries. The GCS scores were used to classify 51 (85%) participants as having mild, moderate, or severe TBI. For nine (15%) participants, injury severity was based on duration of PTA and/or LOC [40]. There were seven mild and two severe designations among these patients.

For the total sample, mean durations of PTA and LOC were, respectively, 743.29 hours (SD = 989.60) and 163.61 hours (SD = 262.14). Only three (5%) individuals were involved in litigation associated with their TBIs.

Procedure

Subsequent to approval by the Institutional Review Boards at the urban trauma centers and a university human subjects committee affiliated with one of the trauma centers, recruitment of participants commenced. All participants were volunteers and signed an appropriate informed consent document. The WAIS-IV was administered to each patient volunteer by a clinical psychology doctoral student or a licensed psychologist with strict adherence to the test manual instructions. All graduate student examinations were conducted under the supervision of a trauma center neuropsychologist while test scoring and research supervision was the responsibility of a graduate psychology program faculty member at the affiliated university. All WAIS-IV administrations were completed between one and 36 months of the date of injury, with all examinees demonstrating stable cognitive functioning at the time of assessment. The figure of 36 months simply describes the range of time (*i.e*., 1-36 months) between injury and WAIS–IV administration in the sample. Prior to testing, stability of cognitive functioning was demonstrated by a single or serial administration of the Galveston Orientation and Amnesia Test [41] with a final score of > 75.

After scoring each protocol, those with FSIQs ≤ 79 were considered to have subnormal intellectual functioning and those with FSIQs ≥ 80 were designated as intellectually normal. The base rates were, respectively, 35% (n = 21) for subnormal ability and 65% (n = 39) for normal functioning. When the same cutoffs for subnormal and normal performance were applied to the GAI, the base rates were 28% (n = 17) and 72% (n = 43). The base rates represent the proportion of individuals out of the total sample of 60 cases who have the condition of interest. If the condition of interest is subnormal intelligence the base rate using the GAI as the cognitive measure is calculated as 17/60 (28%). When the condition of interest is normal intelligence, the base rate is calculated as 43/60 (72%). As mentioned previously, the cutoff of ≤ 79 on the two composites was considered a reasonable estimate of subnormal intelligence because it falls within the Borderline level of functioning and is at or below approximately 92% of the standardization sample.

Based on the standard order of administration specified in the test manual, the following subtest combinations were extracted from each protocol to estimate the FSIQs: two-subtest SF = BD + SI, three-subtest SF = BD + SI + DS, four-subtest SF = BD + SI + DS + MR, and five-subtest SF = BD + SI + DS + MR + VC. After omitting DS, the following subtest combinations were used to estimate the GAIs: two-subtest SF = BD + SI, three-subtest SF = BD + SI + MR, and four-subtest SF = BD + SI + MR + VC. Sums of scaled scores were obtained for each subtest combination and then these sums were transformed into estimated IQs and GAIs with a mean of 100 and standard deviation of 15. These values were obtained using the [27] procedure which is fully explained in Tables B-10 and 11, Appendix B of [42]. It is noted that the estimated IQs and GAIs are obtained in the same manner and are statistically identical. To avoid confusion, the design of this study

mandated that the deviation IQ values used to predict the GAIs be referred to as estimated GAIs.

A simple demographically-based equation was used to derive preinjury FSIQ estimates for each of 56 participants using the Advanced Clinical Solutions for the WAIS-IV and WMS-IV [13]. Four individuals were eliminated due to inadequate information concerning ethnicity. The equation used to obtain IQ estimates was developed on the WAIS-IV standardization sample and education and ethnicity oversample and was based on associations among demographic variables and test performance. The variables used to obtain preinjury IQ estimates were age, education, sex, ethnicity, and geographic region of residence. All cases were coded as "not in the labor force" since occupational information for each participant was unavailable. Each tested FSIQ was compared to the appropriate estimated preinjury IQ. If the tested IQ was lower than the estimated IQ and the magnitude of the difference occurred in < 15% of the combined *ACS* control samples, it was designated as reflecting possible intellectual deterioration. In clinical situations, a discrepancy of this magnitude or larger would likely prompt the clinician to investigate a hypothesis of cognitive loss, although it does not guarantee that mental deterioration has occurred. Smaller differences were considered for the present analysis to be unremarkable and not suggestive of deterioration. In the current sample of TBI patients, the base rates were 45% (n = 25) for possible deterioration and 55% (n = 31) for those without evidence of intellectual deterioration.

Data analysis for the subnormal versus normal intelligence groups and the deteriorated versus not deteriorated groups utilized the classification accuracy statistics of Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value, and Hit Rate. All were derived from formulae provided by [43] and are defined in Table 2. Each statistic can take any value between .00 and 1.00 with interpretation as follows: ≤ .69 = poor, ≥ .70 = moderate, ≥ .80 = strong, and ≥ .90 = excellent. Next, separate ROC analyses were run on the SFIQs and GAI estimates. The purpose was to ascertain how well the ≤ 79 cut-off discriminated between TBI cases with FSIQ-defined normal versus subnormal intelligence. A ROC analysis was also utilized to ascertain the ability of the SFIQs to identify patients with and without suggested intellectual deterioration. Finally, the total sample was ordered according to injury severity on the FSIQ, GAI, and seven SF scores. A series of one-way ANOVAs was conducted to determine if test scores differed significantly across the mild, moderate, and severe injury categories. Also, difference scores were obtained between the preinjury IQ estimates and the FSIQ, GAI, and seven SFIQs. A series of one-way ANOVAs was conducted to determine if the difference scores differed significantly across the mild, moderate, and severe injury categories.

Results

The distributions of scores for the FSIQ, GAI, all SFIQs, and the premorbid IQ estimates were considered normal based on the results of a series of one sample Kolmogorov-Smirnov tests (all D-statistics $p > .05$). Table 3 provides means, standard deviations, and ranges for the WAIS-IV indexes, Full Scale IQs, and GAIs. Also given are means and standard deviations for the estimated SFIQs, estimated SF GAIs, and the demographically-based preinjury IQ estimates.

TABLE 2
Classification Accuracy Statistics

Statistic	Definition
Sensitivity (Sn)	The probability that the SF IQ and GAI correctly identify intellectual impairment for the group.
Specificity (Sp)	The probability that the SF IQ and GAI correctly identify the absence of intellectual impairment for the group.
Positive Predictive Value (PPV)	The probability the SF IQ and GAI are accurate when they predict the presence of intellectual impairment for an individual.
Negative Predictive Value (NPV)	The probability that the SF IQ and GAI are accurate when they predict the absence of intellectual impairment for an individual.
Hit Rate (HR)	The percentages of individuals who are accurately classified by the SF IQ and GAI.

TABLE 3
Means, Standard Deviations, and Ranges for WAIS-IV Variables

Variable	M	SD	Range
Full Scale IQ (n = 60)	84.23	15.96	54 – 136
Verbal Comprehension Index	91.57	15.88	61 – 132
Perceptual Organization Index	89.3	15.82	60 – 129
Working Memory Index	85.07	14.73	55 – 133
Processing Speed Index	78.45	17.19	53 – 135
Two-Subtest Short Form	88.89	16.23	57 – 129
Three-Subtest Short Form	86.76	15.96	54 – 132
Four-Subtest Short Form	87.54	16.38	52 – 127
Five-Subtest Short Form	88.38	16.82	53 – 133
General Ability Index (n = 60)	89.18	15.15	58 – 132
Two-Subtest Short Form	88.89	16.23	57 – 129
Three-Subtest Short Form	89.4	15.78	54 – 126
Four-Subtest Short Form	90.27	15.55	57 – 130
Preinjury Estimated IQ (n = 56)	93.55	7.07	75 – 110

The mean FSIQ was in the low average range and at approximately the 14th percentile relative to the standardization sample. Among the indexes, Verbal Comprehension was the highest while Processing Speed was the lowest, a typical pattern for persons with TBI. All four of the SFIQ values overestimated the Full Scale; whereas, two of the three abbreviated GAIs overestimated the standard GAI. The mean of the preinjury IQ estimates was in the average range and at approximately the 34th percentile. The obtained mean FSIQ (for the 56 participants who also had demographically-based preinjury IQ estimates) was significantly lower than the estimated preinjury IQ, $t(55) = 4.94$, $p < .001$, $d = .66$.

Table 4 presents classification accuracy statistics for the SFIQ and SF GAI estimates using the cutoff of ≤ 79. The percentages of correct classifications (Hit Rates) were markedly superior to the sample base rates of 35% for Full Scale (≤ 79) and 28% for GAI (≤ 79). As indicated in Table 5 only the two-subtest SF estimates of the Full Scale were associated with a high rate of false-positive classifications (*i.e*., 15%). For the GAI estimates, the proportion of false-negative errors was also the highest (8.3%) for the two-subtest SF.

Inspection of Table 4 indicates that for estimating FSIQs, the two, three, four, and five subtest SFIQs showed excellent Sensitivity. The two-subtest SF had strong Specificity and Hit Rates while the three, four, and five subtest SFs had excellent Specificity and Hit Rates. These findings indicate that, in terms of group data, the three-, four- and five-subtest SFIQs were remarkably accurate in classifying participants as intellectually normal or intellectually subnormal. For making individual classification decisions, the Positive Predictive Value and Negative Predictive Value statistics need to be examined. The Positive Predictive Values were poor for the two subtest SF but strong for the three, four, and five subtest SFIQs. The Negative Predictive Values were excellent for each of the SFs. In terms of diagnostic accuracy, these values suggest that the three, four and five subtest SFs were the most useful. In practical terms, if using the four-subtest SF 95% (Sensitivity = 18/19) of participants with subnormal intelligence were correctly identified as were 93% (Specificity = 38/41) of those with normal intelligence. The Positive Predictive Value means that for the individual case the probability is 86% (18/21) that a SF IQ ≤ 79 reflects intellectual sub-normality; whereas, the Negative Predictive Value means that the probability is 97% (38/39) that a SFIQ ≥ 80 reflects normal cognition.

In terms of group relevant statistics, the estimated GAI scores showed moderate Sensitivity when derived from the two-subtest SF; whereas, the three and four-subtest SFs had excellent Sensitivity. The Specificity and Hit Rates for the two, three, and four subtest GAI SFs were excellent as were the Positive Predicted Values and Negative Predictive Values. Overall, the most accurate estimated GAIs were produced by the four-subtest SF. In practical terms, 94% (Sensitivity = 16/17) of the sample with subnormal intelligence was correctly identified as were 100% (Specificity = 43/43) of those with normal intelligence. The Positive Predictive Value means that for the individual case the probability was 100% (16/16) that a SFIQ ≤ 79 reflected intellectual disability; whereas, for the Negative Predictive Value the probability was 98% (43/44) that a SFIQ ≥ 80 reflected normal intelligence.

TABLE 4
Diagnostic Characteristics of Short Form IQs, General Ability Index Estimates, and Preinjury Estimated IQs

Score	Sn	Sp	PPV	NPV	HR
Full Scale IQ					
BD + SI	.92	.81	.57	.97	.83
BD + SI + DS	.9	.92	.86	.95	.92
BD + SI + DS + MR	.95	.93	.86	.97	.93
BD + SI + DS + MR + VC	.95	.93	.86	.97	.93
General Ability Index					
BD + SI	.71	.98	.92	.9	.92
BD + SI + MR	.94	.98	.94	.98	.97
BD + SI + MR + VC	.94	1.00	1.00	.98	.98
Preinjury IQ					
BD + SI	.44	.97	.92	.68	.73
BD + SI + DS	.64	.93	.89	.76	.8
BD + SI + DS + MR	.6	.93	.88	.74	.79
BD + SI + DS + MR + VC	.6	.93	.88	.74	.79

BD = Block Design; SI = Similarities; DS = Digit Span; MR = Matrix Reasoning; VC =Vocabulary

TABLE 5
False-Positive and False-Negative Error Rates for Short Form IQs, General Ability Index Estimates, and Preinjury Estimated IQs

Score	False-Positives	False-Negatives
Full Scale IQ		
Two Subtests	15.0% (9/60)	1.7% (1/60)
Three Subtests	5.0% (3/60)	3.2% (2/60)
Four Subtests	5.0% (3/60)	1.7% (1/60)
Five Subtests	5.0% (3/60)	1.7% (1/60)
General Ability Index		
Two Subtests	1.7% (1/60)	8.3% (5/60)
Three Subtests	1.7% (1/60)	1.7% (1/60)
Four Subtests	0.0% (0/60)	1.7% (1/60)
Preinjury IQ		
Two Subtests	1.8% (1/56)	25.0% (14/56)
Three Subtests	3.6% (2/56)	16.1% (9/56)
Four Subtests	3.6% (2/56)	17.8% (10/56)
Five Subtests	3.6% (2/56)	17.8% (10/56)

A series of ROC analyses were conducted to evaluate the predictive performance of the two, three, four, and five subtest SFIQs and the two, three, and four-subtest GAI estimates. Each ROC analysis included SFIQ as the test variable and normal versus subnormal intellectual functioning as indicated by FSIQ or GAI as the state variable. As can be seen in Table 6, the Area Under the Curve ranged from .956 to .993 when the SFIQ served as the predictor. When the SFs predicted GAI, the Area Under the Curve range was .978 to .997. These results indicate that the SFs used in the present investigation possess excellent ability to discriminate between individuals with normal or subnormal intellectual functioning regardless of whether the FSIQ or GAI serve as the criterion.

Table 4 indicates that all SFIQ estimates of the FSIQ demonstrated poor Sensitivity but excellent Specificity for identifying cases with possible intellectual deterioration. Hit Rate accuracy was strong for the three subtest SF and moderate for the two, four, and five subtest SFIQs. These findings indicate that, in terms of group data, the three, four, and five subtest SFIQs have definite utility for detecting cognitive loss and are very unlikely to label a patient as deteriorated when evidence is lacking. The Positive Predictive Values were excellent for the two-subtest SF and strong for the remaining three SFs. The Negative Predictive Values were poor for the two subtest SF and moderate for the three, four, and five subtest SFs. In practical terms, the three, four and five subtest SFs appear to be the most useful. With the three subtest SF 64% (16/25) of cases with possible intellectual deterioration were correctly identified as were 93% (29/31) of those without such evidence. For the four and five subtest SFs, 60% (15/24) of the sample with possible intellectual deterioration was correctly identified as was 93% (29/31) of those with normal cognitive functioning. The Positive Predictive Values mean that for the individual, the probability was 88% (15/17) that a SFIQ $\leq$ 79 echoed intellectual disability; whereas the Negative Predictive Values indicated that the probability was 74% (29/39) that a SFIQ $\geq$ 80 reflected intact intellectual status. It is noted that false-positive errors were low when the SFIQ estimates were derived from the two, three, four, and five subtest SFs. However, a substantial number of patients with possible intellectual decline were not detected. As Table 5 indicates, the two-subtest SF was associated with the highest rate of false-negatives (25%) while the three-subtest SF produced the lowest, albeit a substantial, proportion of false-negatives (*i.e.*, 16.1%).

A series of ROC analyses were conducted to evaluate the predictive-performance of the two, three, four, and five subtest SFIQs for identifying participants with or without possible intellectual deterioration. As Table 6 indicates, the Area Under the Curve ranged from .845 to .861, indicating that the SFs used in the present investigation possess a strong ability to discriminate between individuals exhibiting possible cognitive loss following TBI and patients without such evidence.

Finally, the influence of injury severity on FSIQ, GAI, and SFIQ performance was examined. Table 7 presents each of the nine WAIS-IV derived scores grouped according to mild, moderate, and severe injury categories determined by GCS results (n = 51) or length of PTA or LOC (n =9). In seven instances, the pattern of scores steadily declined from mild to severe, while in two instances, both involving

TABLE 6
Area Under the Curve for Short Form IQs, General Ability Index Estimates, and Preinjury

Estimated IQs		95% Confidence Interval	
Score	**AUC**	**Lower Bound**	**Upper Bound**
Full Scale IQ			
Two-Subtests	0.956	0.907	1
Three-Subtests	0.973	0.939	1
Four-Subtests	0.991	0.976	1
Five-Subtests	0.993	0.979	1
General Ability Index			
Two-Subtests	0.978	0.95	1
Three-Subtests	0.984	0.959	1
Four-Subtests	0.997	0.99	1
Preinjury IQ			
Two-Subtests	0.849	0.745	0.953
Three-Subtests	0.857	0.758	0.956
Four-Subtests	0.861	0.759	0.963
Five-Subtests	0.845	0.737	0.953

AUC = Area Under the Curve

TABLE 7
Comparisons of Full-Scale IQ, General Ability Index, and Short Form IQs according to Injury Severity Categories

Score	Severity	M	SD	Range	F (2, 59)
Full Scale IQ	Mild	88.39	20.43	54 – 136	1.33
	Moderate	83.11	14.31	65 – 107	
	Severe	81.18	11.44	56 – 109	
Two-Subtest IQ	Mild	90.94	20.15	56.5 – 129.0	0.54
	Moderate	90.98	7.87	79.7 – 102.9	
	Severe	86.53	14.69	59.4 – 120.3	
Three-Subtest IQ	Mild	88.67	19.82	53.8 – 131.5	0.38
	Moderate	87.87	11.87	74.8 – 110.5	
	Severe	84.84	13.68	58.0 – 118.9	
Four-Subtest IQ	Mild	89.9	19.46	52.0 – 127.2	0.44
	Moderate	87.73	14.81	59.6 – 114.4	
	Severe	85.54	14.28	56.8 – 119.2	
Five-Subtest IQ	Mild	90.68	20.42	53.2 – 132.5	0.38
	Moderate	88.3	15.06	71.4 – 116.9	
	Severe	86.52	14.46	55.8 – 116.9	
GAI	Mild	91.13	17.94	58 – 132	0.47
	Moderate	90.56	14.08	72 – 117	
	Severe	87.14	13.13	59 – 112	
Two-Subtest GAI	Mild	90.94	20.15	56.5 – 129.0	0.54
	Moderate	90.98	7.87	79.7 – 102.9	
	Severe	86.53	14.69	59.4 – 120.3	
Three-Subtest GAI	Mild	92.17	18.37	54.0 – 126.0	0.7
	Moderate	90	13.04	72.0 – 111.0	
	Severe	86.93	14.34	58.0 – 120.0	
Four-Subtest GAI	Mild	92.69	19.04	56.5 – 130.0	0.49
	Moderate	90.17	13.4	73.0 – 113.5	
	Severe	88.32	13.07	58.0 – 116.5	

GAI = General Ability Index

the two-subtest SF, this result did not emerge as the highest scores were earned by patients in the moderately injured groups. A one-way ANOVA was conducted for each of the nine types of scores without a single *F*-test reaching statistical significance. Finally, difference scores were calculated separately between the FSIQ, GAI, and SFIQs and their respective preinjury IQ estimates. Results were arranged according to severity groupings and one-way ANOVAs were conducted. All 27 discrepancy scores were negative, indicating that the preinjury estimates were larger than the actual scores and in seven instances there was a steady increase in the size of the discrepancy scores as one moved from the mild to severe injury classification. Discrepancy scores involving the two-subtest SF suggested that patients with moderately severe injuries showed the smallest decline from preinjury estimates. A series of one-way ANOVAs indicated that none of the *F*-tests achieved statistical significance.

Discussion

This present investigation assessed the clinical utility of sequentially administered WAIS-IV SFs to differentiate the intellectual functioning of patients with TBI as either grossly normal (≥ 80) or subnormal (≤ 79) and to determine if the SFs were helpful in identifying the possibility of intellectual deterioration. Following the prescribed order of subtest administration, four SF combinations were created to estimate the FSIQ and, after eliminating the DS subtest, three SFs were identified to predict the GAI (see Table 4). The Advanced Clinical Solutions for the WAIS-IV and WMS-IV was utilized to identify patients with possible intellectual deterioration relative to the obtained FSIQ.

Classification accuracy statistics indicated that the majority of the WAIS-IV SFs were effective for discriminating the intellectual levels of patients with TBI. In settings with a base rate for subnormal intelligence similar to that of the present study, the low false-positive rates of the SFs support their use as screening tools to detect the presence or absence of low-level intellectual functioning. This conclusion applies to both SFIQs and estimated GAIs when using the cutoff of ≤ 79. The SFs were less sensitive to the presence of cognitive decline than they were to the incidence of subnormal intellectual functioning. Nevertheless, the abbreviated forms demonstrated an above chance ability to separate individuals exhibiting possible cognitive loss following TBI from patients without such evidence. With a total sample base rate for cognitive decline of 45% (25/56), the three-subtest SF was able to correctly identify 64% (16/25) of these cases while the four- and five-subtest SFs did so in 60% (15/25) of the sample. The three SFs also appropriately classified 93% (29/31) of those with normal cognitive functioning.

Of the WAIS-IV abbreviations used to produce SFIQ estimates of the Full Scale, the combination of BD + SI was not recommended because it produced a poor Positive Predictive Value and a 15% false-positive error rate. On the other hand, the four-subtest combination of BD + SI + DS + MR appears to be the most useful since it achieved excellent Sensitivity, Specificity, Negative Predictive Value, and Hit Rate along with a strong Positive Predictive Value. Moreover, its classification accuracy statistics, false-positive and false-negative error rates, and ROC outcome were virtually identical to those of the five-subtest SF. The four-subtest SF also requires significantly less time to administer than the five-subtest SF because it does not include the VC subtest. Administration time requirements for individual subtests

are not available for the WAIS-IV, but this information has been published for the WAIS-III. Based on a diagnostically mixed clinical sample, the average time to complete VC was approximately 14 minutes with a range from six to 25 minutes [16]. Until proven otherwise, it seems reasonable to assume that the WAIS-IV VC administration time will be similar.

For estimating GAI scores, the BD + SI combination was not recommended because it had only moderate Sensitivity to intellectual deficit and a relatively high false-negative error rate of 8.3%. The four-subtest SF of BD + SI + MR + VC had excellent classification accuracy statistics, a low false-positive error-rate, and an Area Under the Curve of .997. However, because it includes the VC subtest it will likely require substantial time to administer, suggesting that the three-subtest SF might be a better choice when time is crucial and an estimated GAI is desired. The combination of BD + SI + MR had excellent Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value, and Hit Rate. Moreover, the false-positive and false-negative error rates were < 2% and the Area Under the Curve was .984. These statistics are comparable to those of the four-subtest SF.

Only one previous study examined the accuracy of sequentially administered two-, three, four-, and five-subtest SFs to discriminate normal from subnormal intellectual functioning [10]. Classification accuracy statistics, false-positive and false-negative error rates, and ROC analyses indicated that the two-subtest SF (*i.e.*, BD + SI) was inadequate for assigning individuals to the specified categories. Findings from the present sample of patients with TBI are consistent with those of [10] in that the two-subtest SF was not recommended as a screening instrument, albeit for different reasons. When using the ≤ 79 FSIQ cutoff, the two-subtest SF had poor Sensitivity and a 13% false negative rate in the Ryan et al. study; whereas, the present investigation found that the BD + SI combination was characterized by a poor Positive Predicted Value and a 15% rate of false-positive errors. When assessing patients with TBI, the operating characteristics and false-positive and false-negative rates are highly similar for the three-, four- and five-subtest SFs, suggesting that selection of an abbreviated examination may be determined by the level of patient fatigue or externally imposed time limitations, not just the clinical utility statistics of the SF. These conclusions are consistent with those of Ryan et al.

The conclusion that BD + SI combination is ineffective as a screening device is reinforced by prior research using diagnostically heterogeneous patient samples. [26] employed this SF to estimate FSIQs and GAIs and reported that approximately 20% of the SFIQs and 16% of the estimated GAIs were at least 10-points above or below their target scores. Similarly, [25] found that over 20% of their sample produced SFIQ estimates that were at least ± 10-points away from the actual FSIQs. The current findings, especially when coupled with evidence that the two-subtest SF yields highly inaccurate IQ and GAI estimates, carry practical implications. When practitioners use the sequential administration method for WAIS-IV screening of TBI patients, whenever possible a minimum of three subtests (*i.e.*, BD + SI + DS) should be completed before a decision is made to terminate testing.

The abbreviated GAIs performed in a manner similar to the SFIQs used to estimate the Full Scale. The two-subtest combination was clearly less effective for identifying normal and subnormal intellectual functioning compared to the three- and four-subtest groupings. The two-subtest SF had moderate Sensitivity and

produced 8.3% false-negative errors. However, the remaining SFs had excellent Sensitivity, Specificity, Positive Predictive Value, Negative Predicted Value, and Hit Rates. False-positive and false-negative error rates were < 2% and the Area Under the Curve values were impressively high (*i.e.*, .984 and .997). Based on these findings and the fact that the four-subtest combination may require substantial administration time due to the inclusion of the VC subtest, it is recommended that at least three subtests be completed prior to terminating the abbreviated scale.

A unique aspect of this investigation is the use of WAIS-IV SFs to identify individuals with possible intellectual deterioration. In this regard, the two-subtest SF was the least useful because it had poor Sensitivity and Negative Predicted Value along with a high false-negative error rate. However, in terms of group data, the three-, four-, and five-subtest SFIQs have definite utility for detecting cognitive loss. They are also unlikely to label individuals as experiencing intellectual decline following a TBI when in fact this is not the case. In the present sample, when all SFIQ estimates ≤ 79 were considered suggestive of cognitive deterioration the accuracy rate was 64% for the three-subtest SFIQ and 60% using the four- and five-subtest SFs. However, accuracy was improved by (a) comparing each demographically-based preinjury IQ estimate with the appropriate SFIQ and (b) defining possible intellectual deterioration when the former was 10-points or more above the latter. This yielded a correct classification for possible intellectual decline of 76% (19/25) using the three, four, or five subtest SF.

These findings suggest a need for additional research concerning the utility of SFs for identifying measurable cognitive decline. To improve on the current findings, it is suggested that the demographically-based equation be combined with a word recognition measure, such as the Test of Premorbid Functioning (TOPF) from the Advanced Clinical Solutions [13] software program. This procedure offers more accurate preinjury IQ estimates than those used in the present study for comparison with SFIQs. The ACS provides complex prediction equations that combine demographics, TOPF score, and examinee personal (*e.g.*, current neighborhood wealth) and developmental (*e.g.*, quality of examinee's elementary school education) information. It would also be ideal if the publisher of the WAIS-IV would derive regression equations to predict preinjury intellectual functioning for a variety of SFs, including those evaluated in the present investigation. It is also noted the present study was unable to include patient occupation in the demographic equation. It is unknown who this omission effected the accuracy of the estimated preinjury IQs.

Past studies dealing with the development of WAIS-IV SFs have largely been focused on identifying which combination of subtests produces the most accurate IQ estimates. However, traditional psychometric methods have indicated that it is virtually impossible to identify such subtest groupings that reduce administration time while simultaneously yielding composite scores equivalent to those from the complete scale. These observations, results of [10], and the current study argue that future SF research could benefit from less emphasis on group statistics and more on making diagnostic decisions. When employing a WAIS-IV SF as a screening device it is essential to ascertain the meaning of a given test score for the individual patient as opposed to how well the test score discriminates among diagnostic groups. It seems that knowledge of the Positive Predictive Value (probability of

above average ability) and Negative Predictive Value (probability of lower ability) of a SF is more informative than a research-based assertion that the SFIQ falls within of ± 5 points of the FSIQ. Consider a situation with a cutoff ≥ 110 and Positive Predictive Value and Negative Predictive Value of .50 and .98, respectively. The SFIQ of 108 is below the cutoff for above average ability and the Positive Predictive Value conveys that the likelihood for the condition of interest is at a chance level. However, the excellent Negative Predictive Value indicates that the probability is extremely high that the examinee does not have above average ability. The estimated IQ on the other hand indicates only that the chances are approximately 95% that the FSIQ falls between 103 and 113. A final decision concerning the presence of above average ability represents an educated guess.

This study focused on the effectiveness of sequentially administered SFs to discriminate between TBI patients with either grossly intact or subnormal general ability as well as those with and without possible cognitive deterioration. This approach may prove useful in situations that do not involve personal injury litigation, disability determination, or require a precise FSIQ. For instance, if an examinee scores poorly (*e.g.*, scaled scores = 6; SFIQ = 74) on each of the first four subtests (BD + SI + DS + MR), an examiner may be reasonably confident (*i.e.*, Positive Predictive Value = .86; Hit Rate = .93) that subnormal intelligence is present. In this case the three-subtest SF of the GAI is 76, a finding that reinforces initial conclusion (*i.e.*, Positive Predictive Value = .94; Hit Rate = .97). Moreover, this approach also informs the examiner concerning the possibility that intellectual decline is part of the clinical presentation. Thus, if the four-subtest SFIQ is 10 or more points below the demographically-based preinjury IQ estimate, the possibility of cognitive decline is suggested. These estimated values should prompt administration of the remaining six subtests and the need for a more comprehensive assessment. Conversely, sequential administration of the four subtests may indicate that further intelligence testing is unnecessary and, if desired, the time saved may be used to formally evaluate the individual's memory, language, and/or personality status. In appropriate situations, it may be helpful to interpret a SFIQ or estimated GAI ≥ 90 according to Wechsler's seven category classification system. For example, if an examinee achieves a SFIQ of 125 it might be helpful to report the results as inconsistent with impairment and suggestive of at least average to above average psychometric intelligence.

This investigation demonstrated that sequentially administered WAIS-IV SFs can be useful when testing time is limited and the identification of an examinee as either normal or subnormal intellectually will suffice. When predicting FSIQs, the false-positive and false-negative error rates were relatively low for the three- (5% and 3.3%), four- (5% and 1.7%), and five- (5% and 1.7%) subtest SFs. The situation was even better for the estimated GAI scores in terms of false-positive and false-negative error rates. For example, the false positive and false negative error rates were 0% and 1.7%, respectively, when relying on estimated GAIs derived from a four-subtest SF. These findings were not expected since a relatively low base rate (35%) of the condition of interest (FSIQ ≤ 79) is typically associated with false positive errors [44]. These error rates are comparable to those reported by [10] for their diagnostically heterogeneous sample, but the possibility exists that similar figures may not emerge with other TBI groups. In the current study, 83% (39/47) of participants had abnormal CT scans and 66% (32/47) had GCS scores indicating moderate to

severe injury. Perhaps in a sample composed exclusively of individuals with mild injuries, the utility of the various SFs would be less impressive. However, we can only speculate on this possibility as Table 7 indicates that injury severity had little differential influence on the magnitude of mean scores on the FSIQ, GAI, and their respective SF combinations. It is also important to recognize that Sensitivity and Specificity values change when different cutoff scores are selected (*e.g.*, ≤ 69 versus ≥ 70). Likewise, if different base rates (*e.g.*, 45% versus 15%) for the disorder of interest are encountered there will be increases or decreases in the PPV, NPV, and HR percentages [45]

One possible limitation of the study is a failure to preserve the content coverage of the four factors underlying the scale [46]. When estimating the FSIQ, adherence to the manual's prescribed order of administration yielded three SFs that contained at least one subtest representing the Verbal Comprehension, Perceptual Reasoning, and Working Memory indexes. The SFs developed to predict GAIs represented only two of the indexes. CD and SS were omitted since they are the seventh and tenth subtests in the order of administration. Thus, the Processing Speed construct did not contribute to any of the SFs under investigation. It is acknowledged that TBI results in decreased WAIS-IV scores with the most pronounced changes occurring on CD and SS and the Processing Speed Index [47] [12] [48]. However, it is argued that the omission of these subtests had little or no effect on the results because precise SF IQs or GAI estimations were not the objective; instead, only a dichotomous classification decision was required. Had the focus been on accurate IQ prediction, failure of the SFs to assess processing speed would have certainly introduced a strong bias for overestimation of the traditional scores. This problem is not relevant for GAI estimation since the original score does not include the CD and SS subtests. This view seems reasonable since the SFs created to predict Full Scale IQs all overestimated the composite with an average of 3.66 points. Conversely, the abbreviated GAI means overestimated the standard GAI average by only 0.53 points.

The most significant limitation of the study was the small sample size which may reduce the generalizability of the findings. Another drawback due to the small number of patients is the failure to conduct a more elaborate analysis concerning the influence of injury severity on SF accuracy rates. The importance of including injury severity levels in future SF studies was demonstrated by [47] using a complete administration of the WAIS-IV. Different patterns and levels of performance were seen when controls and patients with mild, moderate, or severe TBI completed the WAIS-IV. Similar variability is likely to emerge when SFs are substituted for the traditional scale. Finally, it may be argued that a limitation of this study is that the BD subtest was included in every SF. We acknowledge that the task may be awkward to administer during bedside assessment and that some patients with impaired upper extremity coordination might experience difficulty arranging the blocks into precise designs. However, this was not a problem for the present investigation because all participants were required to have a functional dominant hand and demonstrate clear consciousness, intact orientation, and the ability to effectively interact with the examiner. It is noted these inclusion criteria may have been unnecessarily rigorous because with standardized administration, as was done in this study, the BD subtest may actually be validly completed using only the nondominant hand [49-50]

References

1. Wechsler D. The measurement and appraisal of adult intelligence (4th edn). Williams & Wilkins. Baltimore, USA. 1958.
2. Golden CJ, Osmon DC, Moses JA and Berg RA. Interpretation of the Halstead-Reitan Neuropsychological Test Battery: A casebook approach. Grune & Stratton, New York, USA. 1981.
3. Reynolds CR, Kamphaus RW. Reynolds Intellectual Assessment Scales and the Reynolds Intellectual Screening Test professional manual. Lutz, FL: Psychological Assessment Resources. 2003.
4. Kaufman AS and Kaufman NL. Kaufman brief intelligence test. (2ndedn), Pearson, San Antonio, Texas, USA. 2004.
5. Wechsler D. WASI-II administration and scoring manual. Pearson, San Antonio, USA. 2011.
6. Wechsler D. WAIS-IV administration and scoring manual. Pearson, San Antonio, USA. 2008a
7. Axelrod BN. Validity of the wechsler abbreviated scale of intelligence and other very short forms of estimating intellectual functioning. Assessment. 2002;9:17-23.
8. Wechsler D. WASI administration and scoring manual. Pearson, San Antonio, USA. 1999.
9. Wechsler D. WAIS-III administration and scoring manual. The Psychological Corporation, San Antonio, USA. 1997.
10. Ryan JJ, Kreiner DS, Gontkovsky ST, et al. Classification accuracy of sequentially administered WAIS-IV short forms. Appl Neuropsychol Adult. 2015;22:409-14.
11. Gaw CE, Zonfrillo MR. Emergency department visits for head trauma in the United States. BMC Emerg Med. 2016;16:1-10.
12. Donders J and Strong CH. Clinical utility of the wechsler adult intelligence scale-fourth edition WAIS-IV after traumatic brain injury. Assessment. 2015;22:17-22.
13. Wechsler D. Advanced clinical solutions for the WAIS-IV and WMS-IV: Administration and scoring manual. Pearson, San Antonio, USA. 2009.
14. Schopp LH, Herrman TD, Johnstone B, et al. Two abbreviated versions of the wechsler adult intelligence scale–III: Validation among persons with traumatic brain injury. Rehabil Psychol. 2001;46:279-87.
15. Rabin LA, Paolillo E, Barr WB. Stability in test-usage practices of clinical neuropsychologists in the United States and Canada over a 10-year period: A follow-up survey of INS and NAN members. Arch Clin Neuropsychol. 2016;31:206-30.
16. Ryan JJ, Lopez SJ, Werth TR. Administration time estimates for WAIS-III subtests, scales, and short forms in a clinical sample. J Psychoeduc Assess. 1998;16:315-23.
17. Kaufman AS. Assessing adolescent and adult intelligence. Allyn & Bacon, Boston, Massachusetts, USA. 1990.
18. Kaufman AS and Lichtenberger EO. Assessing adolescent and adult intelligence. (3rdedn), John Wiley & Sons Inc, Hoboken, New Jersey, USA. 2006.
19. Matarazzo JD. Wechsler's measurement and appraisal of adult intelligence. (5thedn). Williams & Wilkins. Baltimore, Marylnd, USA. 1972.
20. Wechsler D. Measurement of adult intelligence. Williams & Wilkins, Baltimore, USA. 1939.
21. Wechsler D. Manual for the Wechsler Adult Intelligence Scale (WAIS). The Psychological Corporation, San Antonio, USA. 1955.
22. Wechsler D. Manual for the Wechsler Adult Intelligence Scale-Revised (WAIS-R). Harcourt Brace jovanovich San Antonio, USA. 1981.
23. Chen H, Hua M-S. Selecting tetradic short forms of the Taiwan Wechsler Adult Intelligence Scale IV. WAIS-IV Assessment. 2019;26:1633-44.
24. Chen H, Hua M, Zhu J, et al. Selection of factor-based WAIS-III tetrads in the taiwan standardization sample: a guide to clinical practice. Chin J Clin Psychol. 2008;50:91-109.
25. Girard TA, Axelrod BN, Patel R, et al. Wechsler Adult Intelligence Scale-IV dyads for estimating global intelligence. Assessment. 2015;22:441-8.
26. Denney DA, Ringe WK, Lacritz LH. Dyadic short forms of the wechsler adult intelligence scale-IV. WAIS-IV Arch Clin Neuropsychol. 2015;30:404-12.

27. Tellegen A, Briggs PF. Old wine in new skins: Grouping wechsler subtests into new scales. J Consult Psychol. 1967;31:499-506.

28. Wechsler D. WAIS-IV: Technical and interpretative manual. Pearson, San Antonio, USA. 2008b

29. Van Ool JS, Hurks PPM, Snoeijen-Schouwenaars FM., et al. Accuracy of WISC-III and WAIS-IV short forms in patients with neurological disorders. Dev Neurorehabil. 2018;21:101-7.

30. Fan H-Z, Zhu J-J, Wang J, et al. Four-subtest index-based short form of WAIS-IV: psychometric properties and clinical utility. Arch Clin Neuropsychol. 2019;34:81-8.

31. Meyers JE, Zellinger MM., Kockler T, et al. A validated seven-subtest short form for the WAIS-IV. Appl Neuropsychol Adul. 2013;20:249-56.

32. Ward LC. Prediction of Verbal, Performance, and Full-Scale IQs from seven subtest of the WAIS-R. J Clin Psychol. 1990;46:436-40.

33. Ryan JJ, Ward LC. Validity, reliability, and standard errors of measurement for 8two 7-subtest short forms of the Wechsler Adult Intelligence Scale-III. Psychol Assess. 1999;11:207-11.

34. Bulzacka E, Meyers JE, Boyer L, et al. WAIS-IV seven-subtest short form: validity and clinical use in schizophrenia. Arch Clin Neuropsychol. 2016;31:915-25.

35. Lindau M., Mats N. Cross-cultural applicability and reduction of the american seven-subtest shot form of the WAIS on a swedish non-clinical sample. Nord Psychol. 2019;71:148-63.

36. Ryan JJ, Umfleet LG, Gontkovsky ST. Prorating WAIS-IV summary scores for patients with relapsing-remitting multiple sclerosis. Int J Neurosci. 2016;126:1025-9.

37. Lovell M, Franzen MD. Neuropsychological assessment. In: Silver JM, Yudofsky SC and Hales RE (eds), Neuropsychiatry of traumatic brain injury. American Psychiatric Press. Washington DC, USA. 1994; pp.133-60.

38. Kaplan E, Fein D, Morris R and Delis D. WAIS-R as a neuropsychological instrument. San Antonio TX: The Psychological Corporation, California, USA. 1991.

39. Jennett B and Teasdale G. Management of head injuries. Philadelphia, PA: F. A. Davis. 1981.

40. Rao V, Lyketos C. Neuropsychiatric sequelae of traumatic brain injury, Psychosomatics. 2000;41:95-103.

41. Levin HS, O'Donnell VM, Grossman RG. The galveston orientation and amnesia test: A practical scale to assess cognition after head injury. J Nerv Ment Dis. 1979;167:675-84.

42. Sattler JM and Ryan JJ. Assessment with the WAIS-IV. Jerome M Sattler, San Diego, USA. 2009.

43. Smith GE, Cerhan JH and Ivnik RJ. Diagnostic validity. In D. Tulsky D, Saklofske G, Chelune R, Heaton R, Ivnik R, Bornstein R, Prifitera A, Ledbetter M (Eds.). Clinical interpretation of the WAIS-III and WMS-III. Academic Press New York, USA. 2003; pp.273-301.

44. Streiner DL. Diagnosing tests: Using and misusing diagnostic and screening tests. J Pers Assess. 2003;81:209-19.

45. Smith GE, Ivnik RJ and Lucas J. Assessment techniques: Tests, test batteries, norms, and methodological approaches. In: Morgan JE, Ricker JH (Eds.), Textbook of clinical neuropsychology. Psychology Press, New York, USA. 2008; pp.38-57.

46. Smith GT, McCarthy DM, Anderson KG. On the sins of short-form development. Psychol Assess. 2000;12:102-11.

47. Carlozzi NE, Kirsch NL, Kisala PA, et al. An examination of the wechsler adult intelligence scales, fourth edition (WAIS-IV) in individuals with complicated mild, moderate and severe traumatic brain injury (TBI). The Clin Neuropsychol. 2015;29:21-37.

48. Iverson GL, Holdnack JA, Lange RT. Using the WAIS-IV/WMS-IV following moderate-severe traumatic brain injury. In: Holdnack JA, Drozdick LW, Weiss LG, Iverson GL (eds), WAIS-IV, WMS-IV, and ACS: Advanced clinical interpretation. (1stedn) Elsevier Academic Press, New York, USA. 2013; pp.485-544.

49. Briggs PF. The validity of WAIS Performance subtests completed with one hand. Journal of Clinical Psychology. 1960;16:318-19.

50. Ryan JJ, Tree HA. Validity of WAIS–III Performance scale subtests completed with the non-dominant hand. Appl Neuropsychol. 2007;14:52-5.