

Rater Effect on the Behavior Rating Inventory of Executive Function-Adult Version in Traumatic Brain Injury^{*}

Wen Pey-Shan Florida International University, Miami, USA Waid-Ebbs J. Kay Department of Veterans Affairs Medical Center, North Florida/South Georgia Veterans Health System, Gainesville, USA

Velozo Craig A. Medical University of South Carolina, Charleston, USA

The purpose of the study was to investigate whether rater effect exists between self and informant-ratings on the Behavior Rating Inventory of Executive Function-Adult version (BRIEF-A) in individuals with moderate to severe traumatic brain injury (TBI). We recruited 90 individuals with moderate/severe TBI and 89 associated informants. Participants completed a series of assessments including the BRIEF-A. Self and associated informant ratings of the BRIEF-A were analyzed for rater effect using the Many-Facets Rasch Model. The adequate fit statistics indicated raters were not erratic; informants were more severe raters as indicated by significant fixed Chi-square (P < 0.01) and higher rater measure of informants on both Behavior Regulation Index (BRI) and Metacognition Index (MI). At the item level, seven items showed tendency of rater anditem interactions (P < 0.05), and two out of the seven items reached the significant rater and item interactions (P < 0.01). At these two items, the informants were more lenient; that is, they rated the patients as having less executive dysfunction than the patients' self-ratings. We suggested when calculating index scores in individuals with moderate to severe TBI, these two items should be excluded. The results of this study offered a guide to clinicians interpreting incongruent results of patient and informant ratings of the BRIEF-A in moderate to severe TBI.

Keywords: brain injuries, executive function, self report

Introduction

Executive function refers to higher order cognitive processing that are required to successfully perform

^{*}Acknowledgements: The data were collected for an NIH funded project entitled "Developing a Computer Adaptive TBI Cognitive Measure" (NIH R-21#5R21HD045869-03). This work was supported in part by a Pre-Doctoral Fellowship Award, from the United States Department of Veterans Affairs Rehabilitation Research and Development Program. The contents of this publication do not represent the views of the US Department of Veterans Affairs or the United States Government. Special thanks to Dr. Seel's research team at Shepherd Center and Flo Singletary at Brooks Health System for recruiting participants.

Wen Pey-Shan, Assistant Professor, Occupational Therapy Department, Florida International University.

Waid-Ebbs J. Kay, Health Science Specialist, Department of Veterans Affairs Medical Center, North Florida/South Georgia Veterans Health System.

Velozo Craig A., Professor and Chair, Division of Occupational Therapy, Medical University of South Carolina.

goal-directed and purposeful tasks (Roth, Isquith, & Gioia, 2005). The frontal lobes are an essential neuroanatomical structure for carrying out executive function and are vulnerable to damage in a severe traumatic brain injury (TBI). As such, executive dysfunction associated with the frontal lobe damage is common among individuals with severe TBI (Graham, 1999; Povlishock, 1993). While it is agreed that executive function is a common challenge in individuals with severe TBI, how to obtain an accurate measure of executive function is still controversial due to the diversity and complexity of the concept.

Performance measures such as Wisconsin Card Sort test and Trail Making test are traditionally used by neuropsychologists to assess executive function (Vriezen & Pigott, 2002) and require rigorous standardized procedures in a laboratory setting. Using the performance measures in the laboratory settings, individuals are assessed in a context that does not reflect the environment in the real world where there is not a set start and stop time and needed materials are not supplied for the patient. Many studies have challenged the ecological validity of neuropsychological tests because they do not detect the fluid process of executive function that occurs in a "real world" environment (Allen & Ruff, 1990; Chaytor, Schmitter-Edgecombe, & Burr, 2006; Gioia, Kenworthy, & Isquith, 2010; Vriezen & Pigott, 2002).

Given the lack of ecological validity in standardized neuropsychological measures, several questionnaires have been developed to capture aspects of executive function that can only be observed in the real world environment. The Behavior Rating Inventory of Executive Function-Adults version (BRIEF-A) was developed to assess adults' everyday executive function through self and informant report (Roth et al., 2005). The items of the BRIEF-A portray everyday behaviors and provide valuable information in conjunction with neuropsychological measures (Gioia et al., 2010). The BRIEF-A was developed after the BRIEF-children version, which is adopted by National Institute of Neurological Disorder and Stroke as a common data element for TBI (National Institute of Neurological Disorder and Stroke, 2012). The BRIEF-A demonstrated sound psychometrics in individuals with moderate to severe TBI including strong reliability (0.94 and 0.96) and separated individuals into five to six ability levels for both indexes (Waid-Ebbs, Wen, Heaton, Donovan, & Velozo, 2012).

Questionnaires such as the BRIEF-A offer a self-rating option on rating the activities required executive functioning. The advantage of self-rating is that self-rating incorporates the affected person's perspective of their functioning. The BRIEF-A also offer an informant version that can be completed by people who take care of or work with the patients. A major concern in using self-report with individuals with TBI is that they have impaired cognition and self-awareness due to their injury (Abreu et al., 2001; Flashman & McAllister, 2002; Hart et al., 2003). Because of impaired self-awareness, individuals with TBI were found to have a tendency to rate themselves more able than their real ability (Abreu et al., 2001; Allen & Ruff, 1990; Leathem, Murphy, & Flett, 1998; Malec, 2004). On the other hand, informants were thought to be more reliable on answering the questions, but studies also showed that informants' ratings were affected by factors such as caregiver's burden, expectation and observational opportunities (Malec, Machulda, & Moessner, 1997). In summary, self-ratings do not simply overestimate ability over an informant-rating in TBI. The discrepancy/congruency between self-ratings and informant ratings are much more complex and studies have showed varied results.

In some TBI studies, patient self-ratings were actually congruent with informant-ratings in areas of neurobehavioral function, emotion distress and social functioning (Chan & Bode, 2008; Kelley et al., 2014; Leathem et al., 1998; Seel, Kreutzer, & Sander, 1997). Other studies showed that patients underestimated their deficits in certain cognitive, emotional, behavior regulation and participation domains (Cusick, Gerhart, &

Mellick, 2000; Leathem et al., 1998; Malec, 2004; Wilson, Donders, & Nguyen, 2011). Inversely, two studies reported patients overestimated their deficits in certain cognition and participation (Cusick et al., 2000; Malec, 2004). Given these contrasting results from previous studies, it becomes apparent that there are many factors contributing to differences in patient and informant ratings in addition to self-awareness. Studies found factors such as specific domains, severity of injury and time of recovery might be related to the differences of the ratings. For example, self-ratings and informant ratings about physical domains were more congruent than non-physical domains; ratings from individuals with mild TBI were more congruent than ratings from acute patients (Allen & Ruff, 1990; Hart, Sherer, Whyte, Polansky, & Novack, 2004; Leathem et al., 1998; Sherer, 1998). Even the specificity of the items used in questionnaires contributes to congruency between self and informant-ratings (Seel et al., 1997; Sherer, 1998).

In addition to the variety of factors that contribute to rater differences, the statistical methods used to investigate rater discrepancy vary from study to study. The majority of studies used classical test theory to detect discrepancy such as Pearson and Spearman correlation coefficient. These methods do not separate systematic measurement error from random measurement error (Baker, Rounds, & Zevon, 2000). As such, these methods may provide misleading results if raters are consistently giving higher or lower ratings. The Many-Facets Rasch Model (MFRM) based on item response theory offers a better method by providing a model with the "rater" in an equation to account for rater severity/leniency as a systematic error. The MFRM evaluates not only examines rater differences, but also the pattern of the ratings, reliability of the pattern, and rating pattern at the item level (Linacre, 2014).

The systematic error caused by different raters, rating the same behavior is called "Rater Effect" (Scullen, Mount, & Goff, 2000). Rater effect, more specifically rater severity/leniency, estimates the tendency that a rater assigns ratings consistently lower/higher than do other raters (Myford & Wolfe, 2004). Previous research showed that rater effect might be measurement specific (physical versus mental domains; general versus specific questions), and patients specific (stage of recovery, severity of the injury, self-awareness). Given the many factors that affect the ratings of different raters, clinicians need guidance on how to interpret differences between patient and informant ratings on the BRIEF-A. The purpose of this study was to investigate whether rater effect existed between self- and informant-ratings at the index-level and item-level on the BRIEF-A in individuals with moderate to severe TBI. Results provided clinical guidelines to better interpret differences between ratings on the BRIEF-A when it is used in individuals with moderate to severe TBI.

Methods

Participants

A total of 179 participants were enrolled in field testing as a part of an NIH funded study to develop a functional cognitive measure. As part of this larger study, the BRIEF-A was administered in-person to 90 individuals with moderate/severe TBI and 89 associated informants that were recruited from three locations: Shands Hospital in Gainesville, Florida; Brooks Rehabilitation Hospital in Jacksonville, Florida and Shepherd Center in Atlanta, Georgia. The participants with moderate/severe TBI enrolled were either enrolled in inoutpatient rehabilitation (N = 47) or were in the chronic stage of recovery at one-year or more post injury (N = 43).

Inclusion criteria for participants with TBI included: (1) diagnosis with moderate/severe TBI; (2) age from 18 to 89 years; (3) capable of signing informed consent; (4) at outpatient or at least one-year post injury; and (5)

native English speaking. We excluded individuals diagnosed with Schizophrenia, other psychotic disorders or mental retardation. All individuals with TBI are referred to as "patients" for the purposes of this manuscript. Inclusion criteria for informants included (1) at least 18 years old; (2) fluent in English; and (3) observing the individual with TBI at least two hours twice a week. Participants diagnosed with TBI were mainly men while their associated informants were mainly women. Among the patients, 80% did not work, and 65% did not drive. Of the Glasgow Coma Scale (GCS) scores available (40%), 92% of participants had a GCS score of 12 or less. Table 1 presents the characteristics of the participants.

Table 1

Demographics

		Patient $(N = 90)$	Informant $(N = 89)$
Age (year)	Mean $\pm SD$	38.18 ± 15.76	49.93 ± 14.59
	Range	18-84	21-88
Gender	Male	63 (70%)	19 (21.3%)
n (%)	Female	27 (30%)	70 (78.7%)
	White	72 (80%)	75 (84.3%)
Ethnicity	African American	10 (11.1%)	10 (11.2%)
n (%)	Hispanic	4 (4.4%)	3 (3.4%)
	Other/Missing	4 (4.4%)	1 (1.1%)
	High school or less	39 (43%)	24 (27%)
Education	Some college/graduate	44 (49%)	58 (65%)
	Post college	7 (8%)	7 (8%)
	Yes	17 (18.9%)	
Work	No	72 (80%)	
n (70)	Missing	1 (1.1%)	
	Yes	27 (30.0%)	
Driving $n(\%)$	No	59 (65.6%)	
<i>n</i> (70)	Missing	4 (4.4%)	
Time after admit to hospital	Outpatient	5.28 ± 7.10	
(month)	1 year post injury	67.34 ± 81.16	
	Dead	0 (0%)	
	Vegetative	0 (0%)	
	Lower Severe Disability	38 (42.2%)	
	Upper Severe Disability	7 (7.8%)	
GOS-E	Lower Moderate Disability	11 (12.2%)	
	Upper Moderate Disability	13 (14.4%)	
	Lower Good Recovery	16 (17.8%)	
	Upper Good Recovery	1 (1.1%)	
	Missing	4 (4.4%)	

Measures

Behavior Rating Inventory of Executive Function-Adult version (BRIEF-A)(Roth et al., 2005). The BRIEF-A was developed to address the lack of ecological validity in neuropsychological tests. As a supplement to neuropsychological tests, the BRIEF-A measures the effect of executive function deficits on everyday

behaviors. BRIEF-A contains 75 items with two versions: self-report and informant report. Both versions have two indexes: Behavioral Regulation Index (BRI) and Metacognition Index (MI). The BRI consists of 30 items making up four scales: inhibit, shift, emotional control and self-monitor; The MI consists of 40 items making up five scales: initiate, working memory, plan and organize, task monitor, and organization of materials. An additional 5 items assess the acceptability of the ratings. A global executive composite score is calculated by combining the two index scores. A frequency rating scale is used to quantify how often the individual has problems with executive function behaviors: 1 = ``Never'', 2 = ``Sometimes'', and 3 = ``Often''. Thus, lower scores indicate better executive function.

Data Analysis

To investigate rater effects between informants and patients, we conducted the Many-Facets Rasch Model analysis (MFRM) with Andrich Rating Scale Model using the Facets software version 3.71 (Linacre, 2015). With our sample size, we have 95% of confidence that our calibrations are stable within ± 0.5 logits (Linacre, 1994). First, we examined for significance rater effect for index scores using the fixed Chi-square, rater measure, rater separation strata, and reliability of rater separation. The fixed Chi-square of < 0.05 determines a significant rater effect at the index level. Rater measure is a logit estimate of rater severity; we set up the program so a higher rater measure represents a more severe rater, a higher item measure represents a more difficult item, and a higher person measure represents a person with higher executive function. Rater separation strata assesses the number of distinct strata of rater severity in the sample; rater separation close to one indicates no rater effect. Reliability of rater separation indicates the reliability of the rater separation, which should be close to zero when there is no rater effect. A significant rater effect can be concluded with a significant fixed Chi-square, the contrast of rater measures, high rater separation strata and high reliability of rater separation.

Rater fit indexes measure the degree of the rating patterns fit the Rasch measurement model (Myford & Wolfe, 2004; 2003). We considered raters are erratic when their ratings do not fit Rasch model. The criteria for a rater misfitting were rater infit and outfit statistics of MnSq > 1.4 and ZStd > 2 (Wright & Linacre, 1994).

Significant rater effect was further analyzed for rater and item interactions indicating unsystematic inconsistency of rating patterns. An example of unsystematic inconsistency is when a rater being lenient on rating most of the items but being more severe on few items, there would be rater and item interactions on the few items. Rater and item interactions could also occur if a rater is lenient on most of the items, but the degree of leniency on few items were significantly different (a lot more lenient or a lot less lenient) than most of the other times (Linacre, 2014; Myford & Wolfe, 2004; 2003). *T*-tests were used to identify significant rater and item interactions is more stringent than 0.05 but less stringent than a Bonferroni adjustment.

Results

Many-Facets Rasch Model showed there were significant rater effects when index scores were used. The fixed chi-square values were statistically significant (P < 0.01) for the both the BRI and MI ($\chi^2 = 80.5$ with 1 degree of freedom and $\chi^2 = 184.5$ with 1 degree of freedom, respectively). The contrasts of rater measure between informants and patients were 0.42 logits for BRI and 0.56 logits for MI, with informants having a higher rater measure. These rater measure scores indicates that informants were more severe raters, who rated

patients having more executive dysfunction than patients rated themselves in both index scores. Rater separation strata were 8.69 for BRI and 13.07 for MI. The reliability of rater separation index was 0.98 for BRI and 0.99 for MI. The high rater separations and high reliability of rater separation indexes indicated that patients and informant were distinct raters on both indexes and the pattern of distinct raters was very consistent. The fit statistics were within adequate fit criteria indicating raters were not erratic (see Table 2). Figure 1 shows the Facets map displaying rater-person-item measures and their relationships in logits.

+ Measr	-Raters	+Patients	-Items	Scale
 5 ·	Rater severity severe on top			(3)
 4 - 	 	. Patient ability better on top .		
 3 ·	 + ·	*. +	-	
 2 - 	 + -	** ** *** ***	BRI items more difficult on top	
 1 · 	 + - 	*** ** **	+ #13(S) #70(S) #50(S)	
 : : :	Informant	****	<pre>#33(B) #51(E) #44(Sh) #61(Sh) #73(Ih) #69(E) #12(E) #33(E) #55(Ih) #28(E) #5(Ih)</pre>	
: * 0 ' : :	: * : : Patient	* *****	: #/2(E) * #1(E) #67(Sh) : #37(S) #42(E) : #64(S) #23(S) : #16 Has trouble sitting still(Ih) #32(Sh) #19(E) #58(Ih)	(2)
: -1 -	: + -	***. . *	: #22(Sh) #8(Sh) #43(Ih) #29(Ih) #57 People say he/she is too emotional(E) #36(Ih)	
 -2 ·	 + ·	 + -		(1)
 Measr +	+ -Raters	* = 2		

Figure 1. Facets map: raters, patients and items of Behavior Regulation Index (Ih = Inhibition scale; E = Emotional control scale; Sh = Shift scale; In = Initiation scale; W = Working memory scale; S = Self monitor).

RATER EFFECT ON THE BEHAVIOR RATING INVENTORY

Infit and Outfit Statistics for Behavior Regulation Inventory (BRI) and Metacognition (MI)									
Index	BRI					MI			
		Infit	(Dutfit	I	nfit	Ou	utfit	
Fit index	MnSq	ZStd	MnSq	Zstd	MnSq	ZStd	MnSq	Zstd	
Patients	1.02	0.80	1.05	1.50	1.06	2.50	1.10	3.60	
Informants	0.98	-0.80	0.98	-0.70	0.94	-2.70	0.93	-3.10	

Table 2Infit and Outfit Statistics for Behavior Regulation Inventory (BRI) and Metacognition (M

In spite of informants being the more severe raters, seven items showed a tendency of rater and items interaction: informants were more lenient as they rated patients less impaired than self-ratings on five items (P < 0.05); inversely, on two items informants rated patient a lot more impaired than the average pattern (see Table 3). Only #16 and #57 reached statistical significant level of P < 0.01.

Table 3

#25 (In)

Informants were lenient raters on: Т Items Patients Informants Contrast Joint SE Prob. #16 Has trouble sitting still (Ih) 0.42 -0.58 1.00 0.26 3.89 < 0.01 #57 People say he/she is too -0.42-1.20 0.78 0.28 2.73 < 0.01 emotional (E) #61 (Sh) 0.45 -0.150.6 0.25 2.37 0.02 #58 (Ih) 0.03 -0.51 0.26 0.04 0.55 2.1 -0.43 0.04 #26 (W) 0.12 0.55 0.26 2.13 Informants were much more severe raters on: Т Р Items Patient informant Contrast Joint SE 0.27 0.03 #8 (Sh) -0.570.02 -0.59 -2.21

Rater and Item Interaction for Behavior Regulation Index and Metacognition Index

-0.41

Notes. Ih = Inhibition scale; E = Emotional control scale; Sh = Shift scale; In = Initiation scale; W = Working memory scale.

-0.62

0.26

-2.37

0.02

0.21

The second and third columns show the item measures rated by patients and informants in logits. For example, item measure of #16 was 0.42 logit when rated by patients, but -0.58 logit when rated by informants. The fourth column, "Contrast", shows the difference of the item measures between two raters. For example, the difference of item measures for item #16 is 0.42-(-0.58) = 1.00.

Discussion

The purpose of the study was to investigate whether a rater effect existed between self-report and informant-report on the BRIEF-A in a moderate to severe TBI sample. We found that informants rated patients with more executive dysfunction than did the patients themselves in index scores. This finding was supported by significant fixed Chi-square, high rater separation strata and high reliability of rater separation, and higher rater measure of informants (Myford & Wolfe, 2004).

Even though informants were more severe raters at the index scores, when looking into the rating of each item, they did not execute this systematic pattern across all the items. Informantsrated patients having more difficulty on most of the items in similar degree, but on five items informants broke this rating pattern and rated patients having less difficulty than patients self-ratings (P < 0.01 on two items). Additionally, on two items, informants rated patients having a lot more difficulties than the average pattern. Seel and colleagues (1997)

found the incongruence of patients and informants ratings on only certain items of the communication scale of the Neuro Functioning Inventory, where caregivers rated patients less impaired than patients' self-ratings. Examples of the items such as "Ringing in ears" "Uncomfortable around others" represent more of inner struggle and are difficult to be observed by caregivers (Seel et al., 1997). Similarly, in our study, informants deviated from the systematic pattern of rating patients more impaired by being more lenient on five items of which two items reached significant level of P < 0.01. These two items are item #16 "Has trouble sitting still" and item #57 "People say he/she is too emotional". Possible reasons for this rater difference might be that outsiders (informants) may have less chance to observe the behaviors that describe an internal struggle. Sitting still might be an easy task in anormal population, but insiders (individuals with TBI) might struggle to inhibit the impulse of moving that is not observable by an outsider. Bias and judgements play a significant role in assigning ratings when the ratings are assigned by outsiders (Wolfson, Doctor, & Burns, 2000). Additionally, informants may have less opportunity to hear what other people say about the patients than did the patients themselves (Malec et al., 1997); this might explain why informants rated patients as having fewer problems on the item "People say he/she is too emotional". The items with highly significant rater and item interaction such as #16 and #57 should be identified and excluded from the scoring in moderate to severe TBI.

This is the first study to examine rater effect of BRIEF-adult version in TBI and one of the few study investigated self versus informant ratings using Rasch model. Willson and colleagues (2011) examined rater agreement in the BRIEF-children version in adolescents with TBI, and found that parents rated adolescents with TBI having more problems on both indexes of the BRIEF (Wilson et al., 2011). However, due to the limitation of their statistical method, they did not examine whether adolescents with TBI rated consistently or erratically, nor did they examine rater differences at the item level.

The advantage of using the MFRM is that the analysis goes beyond correlation and mean difference. Few studies have taken advantage of using Rasch model to identify rater effect. The MFRM applies measurement theory to inspect the rating patterns, and can integrate many potential measurement facets into model, instead of treating everything as measurement error. As such, the MFRM has capacities to provide more information including rater severity, respondents' ability, item difficulty, erratic raters, erratic items, and inconsistent ratings due to rater and item interaction. This model provides a means to identify items and raters that do not follow response pattern for further investigation.

Many studies have examined the concordance of patients' self-ratings and ratings of informants (Kelley et al., 2014; Leathem et al., 1998; Malec, 2004; Seel et al., 1997; Sherer, 1998). Unfortunately, the results varied depending on instruments, the domains of the instruments, how specific or general the items were written, etc.. The majority of these studies concluded that patients with TBI overestimated their abilities due to lack of self-awareness (Godfrey, Partridge, Knight, & Bishara, 1993). This is consistent with our findings as our patients also reported less executive dysfunction at the factor/index level on the BRIEF-A.

The validity of the self-report from individuals who have cognitive deficits such as TBI has been questioned. However, we did not find erratic raters. This result suggests that although patients were more lenient raters at the index scores, their ratings followed the measurement model: individuals with TBI consistently rated more challenging tasks as difficult and less challenging tasks as easy. Therefore, patients self-report provides reliable and valuable information on their perspective of their executive function, even though they might be lenient on their ratings. This finding is important for researchers or clinicians who have difficulty accessing caregivers or only have one rater available. To resolve the difficulty of comparing ratings

from different raters across difference studies and comparing outcomes in longitudinal studies where data are collected from different raters, creating a conversion table with a larger sample will be useful so that the scores can be adjusted for rater differences (Fisher, 1993).

Our study was not without limitations. First, although our sample was identified as moderate to severe TBI, we did not have complete supporting medical record documentation to verify their severity level. Second, since our sample was recruited in the southeast region of the nation, generalizability of our finding is limited to the geographic area and severity of the TBI.

Clinical Implications

Because the informants and patients showed significantly unsystematic response patterns on two of the BRIEF-A items, use of these two items can dilute the accumulated index score. We suggest to excluding these two items to provide a more accurate index score. Users may also isolate these two items and use them to provide additional information. For example, greater differences on these two items may indicate that informants overlook the internal struggles of patients.

Conclusion

Clinicians and researchers have found the BRIEF-A to be a useful tool in examining real world executive functioning (Rabin et al., 2006; Schiehser et al., 2011; Waid-Ebbs et al., 2012). BRIEF-A has demonstrated good psychometrics in a TBI sample and offers two perspectives of the patient's functioning; self and informant (Waid-Ebbs et al., 2012). However, differences between these raters can be difficult for the clinicians and researchers to interpret. Our study has demonstrated the first rater effect analysis in the adult version of the BRIEF. Overall, the informants were more severe raters at the factor/index level and two items were unsystematically rated differently (informant ratings were more lenient). We recommend these two items to be excluded when interpreting the index scores in patients with moderate to severe TBI.

References

- Abreu, B. C., Seale, G., Scheibel, R. S., Huddleston, N., Zhang, L., & Ottenbacher, K. J. (2001). Levels of self-awareness after acute brain injury: How patients' and rehabilitation specialists' perceptions compare. *Archives of Physical Medicine and Rehabilitation*, 82(1), 49-56. doi:S0003-9993(01)95199-8 [pii]
- Allen, C. C., & Ruff, R. M. (1990).Self-rating versus neuropsychological performance of moderate versus severe head-injured patients. *Brain Injury*, 4(1), 7-17.
- Baker, J. G., Rounds, J. B., & Zevon, M. A. (2000). A comparison of graded response and rasch partial credit models with subjective well-being. *Journal of Educational and Behavioral Statistics*, 25(3), 253-270.
- Chan, R. C., & Bode, R. K. (2008). Analysis of patient and proxy ratings on the dysexecutive questionnaire: An application of rasch analysis. *Journal of Neurology, Neurosurgery, and Psychiatry*, *79*(1), 86-88. doi:jnnp.2007.117184[pii]
- Chaytor, N., Schmitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. Archives of Clinical Neuropsychology. *The Official Journal of the National Academy of Neuropsychologists*, 21(3), 217-227. doi:S0887-6177(06)00018-7[pii]
- Cusick, C. P., Gerhart, K. A., & Mellick, D. C. (2000). Participant-proxy reliability in traumatic brain injury outcome research. *The Journal of Head Trauma Rehabilitation*, 15(1), 739-749.
- Fisher, A. G. (1993). The assessment of IADL motor skills: An application of many-faceted rasch analysis. *The American Journal of Occupational Therapy: Official Publication of the American Occupational Therapy Association*, 47(4), 319-329.
- Flashman, L. A., & McAllister, T. W. (2002). Lack of awareness and its impact in traumatic brain injury. *NeuroRehabilitation*, 17(4), 285-296.
- Gioia, G. A., Kenworthy, L., & Isquith, P. K. (2010). Executive function in the real world: BRIEF lessons from mark ylvisaker. *The Journal of Head Trauma Rehabilitation*, 25(6), 433-439. doi:10.1097/HTR.0b013e3181fbc272[doi]

- Godfrey, H. P., Partridge, F. M., Knight, R. G., & Bishara, S. (1993). Course of insight disorder and emotional dysfunction following closed head injury: A controlled cross-sectional follow-up study. *Journal of Clinical and Experimental Neuropsychology*, 15(4), 503-515. doi:10.1080/01688639308402574 [doi]
- Graham, D. (1999). Pathophysiological aspects of injury and mechanisms of recovery. In M. Rosenthal (Ed.), *Rehabilitation of the adult and child with trauamtic brain injury* (3rd ed., pp. 19-41). Philadelphia, P.A.: FA Davis Company.
- Hart, T., Sherer, M., Whyte, J., Polansky, M., & Novack, T. A. (2004). Awareness of behavioral, cognitive, and physical deficits in acute traumatic brain injury. *Archives of Physical Medicine and Rehabilitation*, 85(9), 1450-1456. doi:S0003999304002941[pii]
- Hart, T., Whyte, J., Polansky, M., Millis, S., Hammond, F. M., Sherer, M., ... Kreutzer, J. (2003). Concordance of patient and family report of neurobehavioral symptoms at 1 year after traumatic brain injury. *Archives of Physical Medicine and Rehabilitation*, 84(2), 204-213. doi:10.1053/apmr.2003.50019[doi]
- Kelley, E., Sullivan, C., Loughlin, J. K., Hutson, L., Dahdah, M. N., Long, M. K.,...Poole, J. H. (2014). Self-awareness and neurobehavioral outcomes, 5 years or more after moderate to severe brain injury. *The Journal of Head Trauma Rehabilitation*, 29(2), 147-152. doi:http://dx.doi.org.ezproxy.fiu.edu/10.1097/HTR.0b013e31826db6b9
- Leathem, J. M., Murphy, L. J., & Flett, R. A. (1998). Self- and informant-ratings on the patient competency rating scale in patients with traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology*, 20(5), 694-705. doi:10.1076/jcen.20.5.694.1122[doi]
- Linacre, J. M. (1994). Sample size and item calibration stability. Rasch Measurement Transactions, 7(4), 328.
- Linacre, J. M. (2014). A user's guide to FACETS rasch-model computer programs. Retrieved from www.winsteps.com
- Linacre, J. M. (2015). Facets software (version 3.71)[computer software]. Beaverton, Oregon: Winsteps.com.
- Malec, J. F. (2004). Comparability of mayo-portland adaptability inventory ratings by staff, significant others and people with acquired brain injury. *Brain Injury*, 18(6), 563-575. doi:10.1080/02699050310001646134 [doi]
- Malec, J. F., Machulda, M. M., & Moessner, A. M. (1997). Differing problem perceptions of staff, survivors, and significant others after brain injury. *Journal of Head Trauma Rehabilitation*, 12(3), 1-13.
- Myford, C. M., & Wolfe, E. W. (2004). Detecting and measuring rater effects using many-facet rasch measurement: Part 2. Journal of Applied Measurement, 5(2), 189-227.
- Myford, C. M., & Wolfe, E. W. (2003). Detecting and measuring rater effects using many-facet rasch measurement: Part I. Journal of Applied Measurement, 4(4), 386-422.
- National Institute of Neurological Disorder and Stroke. (2012). Retrieved from https://commondataelements.ninds.nih.gov/#page =Default
- Povlishock, J. T. (1993). Pathobiology of traumatically induced axonal injury in animals and man. *Annals of Emergency Medicine*, 22(6), 980-986. doi:S0196-0644(05)82738-6 [pii]
- Rabin, L. A., Roth, R. M., Isquith, P. K., Wishart, H. A., Nutter-Upham, K. E., Pare, N., ...Saykin, A. J. (2006). Self- and informant reports of executive function on the BRIEF-A in MCI and older adults with cognitive complaints. Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists, 21(7), 721-732. doi:S0887-6177(06)00107-7[pii]
- Roth, R., Isquith, P., & Gioia, G. (2005). Behavior rating inventory of executive function-adult version. Lutz, FL: PAR.
- Schiehser, D. M., Delis, D. C., Filoteo, J. V., Delano-Wood, L., Han, S. D., Jak, A. J., ...Bondi, M. W. (2011). Are self-reported symptoms of executive dysfunction associated with objective executive function performance following mild to moderate traumatic brain injury? *Journal of Clinical and Experimental Neuropsychology*, 33(6), 704-714. doi:10.1080/13803395.2011.553587[doi]
- Scullen, S. E., Mount, M. K., & Goff, M. (2000). Understanding the latent structure of job performance ratings. *The Journal of Applied Psychology*, 85(6), 956-970.
- Seel, R. T., Kreutzer, J. S., & Sander, A. M. (1997). Concordance of patients' and family members' ratings of neurobehavioral functioning after traumatic brain injury. *Archives of Physical Medicine and Rehabilitation*, 78(11), 1254-1259. doi:http://dx.doi.org/10.1016/S0003-9993(97)90340-3
- Sherer, M. (1998). Characteristics of impaired awareness after traumatic brain injury. Journal of the International Neuropsychological Society, 4(04), 380.
- Vriezen, E. R., & Pigott, S. E. (2002). The relationship between parental report on the BRIEF and performance-based measures of executive function in children with moderate to severe traumatic brain injury. *Child Neuropsychology: A Journal on Normal* and Abnormal Development in Childhood and Adolescence, 8(4), 296-303. doi:10.1076/chin.8.4.296.13505[doi]

- Waid-Ebbs, J. K., Wen, P. S., Heaton, S. C., Donovan, N. J., & Velozo, C. (2012). The item level psychometrics of the behaviour rating inventory of executive function-adult (BRIEF-A) in a TBI sample. *Brain Injury*, 26(13-14), 1646-1657. doi:10.3109/02699052.2012.700087[doi]
- Wilson, K. R., Donders, J., & Nguyen, L. (2011). Self and parent ratings of executive functioning after adolescent traumatic brain injury. *Rehabilitation Psychology*, *56*(2), 100-106. doi:10.1037/a0023446[doi]
- Wolfson, A. M., Doctor, J. N., & Burns, S. P. (2000). Clinician judgments of functional outcomes: How bias and perceived accuracy affect rating. Archives of Physical Medicine & Rehabilitation, 81(12), 1567-1574.
- Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. Rasch Measurement Transactions, 8(3), 370.