In press at *Archives of Clinical Neuropsychology*

Evaluating Personality Assessment Inventory (PAI) response patterns in active-duty personnel with head injury using a Latent Class Approach

Paul B. Ingram, PhD
Department of Psychological Sciences
Texas Tech University, Lubbock, TX
Dwight D. Eisenhower Veteran Affairs Medical Center
Eastern Kansas Veteran Healthcare System
Leavenworth, KS
ORCid: 0000-0002-5409-4896

Tristan T. Herring
Department of Psychological Sciences
Texas Tech University, Lubbock, TX
ORCid: 0000-0001-7468-794X

Patrick Armistead-Jehle, PhD, ABPP-CN
Munson Army Health Center, Fort Leavenworth, KS

Author notes. The views, opinions, and/or findings contained in this article are those of the authors and should not be construed as an official Defense Health Agency, Department of Veterans Affairs, or U.S. Government position, policy or decision unless so designated by other official documentation.

Correspondence concerning this manuscript should be addressed to the first author, Dr. Paul Ingram, 2810 18th Street, Box 42051, Lubbock Texas 79409, Texas Tech University, Department of Psychological Sciences. E-mail: Paul.Ingram@ttu.edu Phone (806) 834-3354

Abstract

Objective: Previous research has found that among those with brain injury, individuals have a variety of different potential symptom sets which will be seen on Personality Assessment Inventory (PAI). The number of different groups and what they measure has varied by study.

Method: In Active-Duty personnel with a remote history of mild traumatic brain injury (*n* = 384) who were evaluated at a neuropsychology clinic, we used a retrospective database to examine if there are different groups of individuals who have distinct sets of symptoms as measured on the PAI. We examined the potential of distinct groups of respondents by conducting a latent class analysis of the clinical scales. Post-hoc testing of group structures was conducted on concurrently administered cognitive testing, performance validity tests, and the PAI subscales.

Results: Findings indicate that a pattern of broad symptom severity as the most probable reason for multiple groups of respondents, suggesting that there are not distinct symptom sets observed within this population. Pathology levels were the most elevated on internalizing and thought disorder scales across the various class solutions.

Conclusion: Findings indicate that amongst Active-Duty Service Members with remote brain injury, there are not distinct groups of respondents with different sets of symptom types as has been found in prior work with other neuropsychology samples. We conclude that the groups found are likely a function of general psychopathology present in the population/sample, rather than bone fide differences.

Keywords: PAI, Assessment, Military, TBI, Latent Class Analysis

Evaluating Personality Assessment Inventory (PAI) response patterns in active-duty personnel with head injury using a Latent Class Approach

Cognitive and psychological evaluation after brain injury represents a pressing clinical issue for military psychologists. Approximately 80% of the 350,000 assessments conducted by the Department of Defense between 2000 and 2015 include some evidence of brain injury (Defense and Veteran Brain Injury Center Annual Report [DVBIC], 2016), documenting substantially higher rates than are observed in the general U.S. population (e.g., Frost et al., 2013; Schneider et al., 2018). Within the active-duty military population, an increased rate of concussion has been noted with combat deployments; however, most of these injuries occur when stationed within a given area, rather than while doing combat patrols (e.g., Bass et al., 2012; Rona et al., 2012). Mild traumatic brain injury (mTBI) is the most common form of brain injury experienced by U.S. Service Members (SM) (Terrio et al., 2009) and represents a specific area of concern with respect to military psychological and neuropsychological evaluations. Indeed, studies have found that 15 to 23% of SMs sustained a TBI since joining the military (e.g., Hoge et al., 2008; Ivins et al., 2003; Terrio et al., 2009). Thus, evaluation of brain injured SMs is common for military psychologists, with inaccurate diagnosis potentially resulting in not only inappropriate rendering of health care, but also substantial long-term financial impacts (Denning & Shura, 2019). Given these effects, neuropsychologists working in the military setting often rely on broadband measures of personality to assess psychopathology and symptom validity (Martin et al., 2015; Worthen & Moering, 2011).

With regards to psychological testing, the majority of research on patients who have sustained a mild traumatic brain injury (mTBI) has focused on validity scale performance (e.g., Ingram et al., 2021; Jurick et al., 2019; Meyers et al., 2014; Schroeder et al., 2012; Tarescavage et al., 2013; Thomas & Youngjohn, 2009; Tylicki et al., 2021; Wygant et al., 2011; Youngjohn et al., 2011) or the impact of invalid symptom responding on clinical scale performance (Gervais et al., 2011; Jones et al., 2009, 2012). Despite a strong base of research on validity scales across broadband measures, less work has been done on the clinical scales of these same broadband measures, particularly amongst military and veteran individuals. In the limited research that has been completed, a consistent finding is the relationship between brain injury severity and scales assessing somatic and/or cognitive difficulties (e.g., Arbisi et al., 2011; Bolinger et al., 2014; Goldsworthy & Donders, 2019; Whiteside et al., 2012; Youngjohn et al., 2011). Additionally, those with mTBI produce high rates of invalid responding and elevated clinical scale scores (Jurick et al., 2019). Heighted internalizing distress is also frequently evident in those with mTBI (Whiteside et al., 2012)

Across personality measures, a few studies have looked at those with brain injury using participant grouping methods (e.g., cluster and latent class/profile analysis methods). In general, these studies have produced irregularities and, as such, offer little cohesive guidance on how to interpret results amongst those who have sustained brain injuries. Some of the inconsistencies in findings may reflect, to a degree, the variability of recovery from brain injury (e.g., symptom presentation as a function of time since injury; Iverson, 2005), which differ across studies along with settings in which they were done. The limited number of studies, as well as the variability of findings from those studies, offers little confidence for clinicians tasked with making meaningful profile interpretations. To further the use of broadband measures as diagnostic aids, it must be determined if a consistent and discernable pattern of symptoms can be expected, as well as how long such a pattern is evident for given injury recovery timelines.

Broadly, there are two particularly relevant, but distinct, approaches to grouping response styles on symptom measures that are used to determine patterns of response. The first identifies co-occurring variation amongst instruments scales (i.e., factor analysis) and determines an implied (i.e., latent) symptom structure. The second focuses on finding co-occurring groups of individuals based on co-occurring elevations (cluster/latent class). Exploratory Factor Analysis (EFA) is conceptually orthogonal to the patient grouping methods of latent clustering. Both methods have been used to study the response patterns of those with a history of TBI on broadband personality measures; however, patient clustering methods (i.e., cluster/latent class) are more appropriate for identifying groups of individuals because they treat the person as the unit of analysis, rather than items. Two recent studies (Johnstone et al., 2022 and Goldsworthy & Donders, 2019) represent examples of these distinct methods in samples of patients with a history of TBI.

Using factor analysis, Johnstone et al. (2022) examined response patterns among active-duty service members with a history of TBI evaluated in an outpatient neuropsychologic clinic approximately one year since injury. They found four conceptual clusters of scales. These four latent scale clusters consisted of: (1) general distress on psychiatric, cognitive, and somatic symptoms, (2) interpersonal/social distress symptoms, (3) depression-specific symptom distress, and (4) substance-use related symptoms. Accordingly, Johnstone et al conclude individuals are likely to fall into one of these four groups when they have a brain injury. These clustered groups of scales (described by Johnstone and colleagues as "face valid" clusters) were based on item content of the Personality Assessment Inventory (PAI; Morey, 1991), and thus lacked any concurrent validation (i.e., external criterion) for observed groups as is typical for interpretation of any construct (see Cronbach & Meehl, 1955). Moreover, Johnstone et al failed to meet necessary sampling to conduct their analyses (Brown, 2015) and had more Likert-type items analyzed (k = 214) than overall participants (n = 210). Likewise, on average, scale scores comprising these four areas did not rise to a level which would be indicative of clinical impairment (i.e., mean scores fell well below recommended cut-scores, and many scales [including both substance use scales on the PAI] matched normative means). It could be argued that EFA was used when a participant centered analytic approach would have been more appropriate, given the purpose of EFA for item response pattern rather than for identifying groups of individuals. This concern adds to the other methodological limitations described above (e.g., insufficient sample size, scores indicating normative rather than pathological concern).

In contrast, Goldsworthy and Donders (2019) used a patient-centered, two-stage cluster analytic approach to see how non-military individuals with a brain injury in a rehabilitation hospital (1 to 12 months post-injury) could be grouped on a comparable broadband personality measure MMPI-2-Restructured Form (MMPI-2-RF; see Tellegen & Ben-Porath, 2008/2011 for correlations demonstrating construct overlap). While a distinct instrument from the PAI, the conceptual overlap of scales provides evidence that the MMPI-2-Rf and PAI are meaningful proxies to one another, able to be conceptualized together. Goldsworthy and Donders (2019) identified a four-cluster pattern. Specifically, they concluded that variation existed as a function of (1) symptom severity (cluster one [substantial elevation across most scales] versus cluster four [low symptom endorsement and scale elevation]) and (2) symptom pattern (cluster two [depression and somatic] versus cluster three [somatic only]). Additionally, findings include: (1) cluster one had the mildest injuries, but the highest rate of concurrent compensation seeking and invalid validity scales (60%); (2) cluster four had a higher frequency of severe TBI despite lowest symptom endorsement; and (3) gender differences were evident across clusters two (more male) and three (more female). A major limitation of this study was that invalid responding was not screened prior to interpreting clinical scales, despite invalid responding greatly skewing substantive scale interpretation in those TBI (Jurick et al., 2019). Likewise, model fit was not evaluated, meaning that findings were non-falsifiable and alternative models could not be evaluated for better fit. Others examining mTBI populations symptom patterns on the MMPI-2-RF have found no need for subgroups. Instead, Childs et al. (2022) offered a single comparison group for outpatients 2- and 24-months post injury, with elevated concerns most evident on scales measures somatic and/or cognitive substantive difficulties and across most of the over-reporting scales.

Knowing the potential groups offers a way for clinicians to utilize a nomological framework for this population, potentially aiding in diagnostic and outcome prediction (e.g., Morris et al., 2022) on TBI recovery pattern prediction (Carney et al., 2014; Iverson, 2005). At present, the literature on TBI patient symptom cluster patterns are limited to such a degree that engaging such methodology is not possible, despite similar time since injury and, presumptively, similar levels of impairment (Goldsworthy & Donders, 2019; Johnstone et al., 2022). There is little consensus between these studies. Each study faces their own limitations, such as inappropriate method, failure to remove subjects with invalid responding, limited samples, and incongruent findings. The only consistent response pattern observed is a general distress group, but even this group may reflect etiologies other than brain injury per se (general distress; see Ingram et al., 2022). In short, research is needed to aid in understanding symptom presentation patterns amongst those with a history of TBI using broadband personality measures (Childs et al., 2022). Within the framework of a limited literature base, the current study extends the research on PAI substantive scale presentation in an active-duty military sample with a history of

TBI by using latent class analysis to evaluate scale patterns observed in an outpatient military neuropsychology clinic. Latent class analysis is like cluster analysis but includes fit metrics that enable an objective, empirical basis for the comparison of models (see Windgassen et al., 2018 for a description of various clustering approaches and the contrasting elements of each).

Current Study

We examine PAI responses amongst individuals service members with a documented history of mTBI, as operationally defined by Department of Defense [DoD]/Veteran Affair [VA] criterion, and which were tested within an outpatient neuropsychological clinic. In general, these mTBI injuries are fairly remote in origin (M = 61.6 months since last TBI; SD = 51.8). This study was designed to expand contextual understand about if, and how, remote head injury should be considered to correspond with testing performance and symptom endorsement.

This study tests three hypotheses. First, consistent with prior work on TBI presentation across broadband measures (Goldsworthy & Donders, 2019; Johnstone et al., 2022), we hypothesized the emergence of high/low scoring groups reflecting high distress in some individuals. Second, we anticipated that those with more severe injuries will demonstrate a low symptom pattern and that those who fail performance validity testing will be better represented in the high symptom group (Goldsworthy & Donders, 2019; Jurick et al., 2019; Thomas & Youngjohn, 2009). Our third hypotheses states that specific groups would not be identified (i.e., the clusters identified by Goldsworthy and Donders (2019) which, for instance, demonstrate a depressive/somatic pattern [cluster two] and a substantially lower and somatic only presentation [cluster three]). We formulated the third hypothesis partially as a function of extraction artifacts observed in class/cluster analysis that can result in over-extraction as a function of dimensional symptom severity (e.g., Aguerrevere et al., 2018; Soldino et al., 2019), which has been observed on the PAI (Ingram et al., in press), and because of the sample dependent nature of extracted factors which do not vary from normative samples (e.g., Johnstone et al., 2022). Thus, we hypothesized that clusters would either merge into a single cluster of somatically focused symptoms (e.g., 3 cluster solution), or fold into the high/low symptom clusters entirely, resulting in a more probable two-cluster solution. Clusters identified reflecting normative score patterns may emerge but are not considered meaningful if they are not symptomatically distinct from normative behavior. Our hypothesized findings and analytic strategy avoid potential over pathologizing and offers more meaningful guidance.

Method

Participants

Potential participants were SMs evaluated in an outpatient neuropsychological clinic who had completed psychological testing. We excluded potential participants undergoing Medical Board and Temporary Disability evaluations because of the impact of potential financial gain on response validity and the low number of those participants who would make it difficult to establish a unique ‘class’ of respondents. We also excluded individuals who exceeded recommended cut-scores on primary PAI validity scales (INC$\geq$ 73; INF$\geq$ 75; NIM $\geq$ 110; PIM $\geq$ 68). These exclusion scores were selected for various reasons, including: (1) NIM’s relationship with clinical symptoms (e.g., Ingram et al., 2021; Bellet et al., 2017) and because prior publication with this data has evidenced no notable differences in classification accuracy on NIM based on PVT outcomes (Morris et al., 2022) and (2) contradicting exclusion approaches used in prior cluster analysis studies leading to use of some scales but not all possible measures, allowing some generalization to each of the extant clustering papers conducted previously. Exclusions were not made for non-standard primary scales because of the limited literature on these measures (see McCredie & Morey, 2018). Data on PAI validity scale performance using the participants in this study (including those excluded) has been previously published (Armistead-Jehle et al., 2020; Ingram et al., in press; Morris et al., 2022). The substantive clinical scales have yet to be analyzed.

The study included a final sample of 384 SMs (Table 1), referred to a U.S. Army Health Center outpatient neuropsychology clinic for neuropsychological evaluation. The average age of the sample was 37.4 years (SD = 7.3), with an average education of 15.5 years (SD = 2.3). The sample was predominantly male (*n* = 346, 90.1%), white (*n* = 299, 77.9%), and contained slightly more officers (*n* = 208, 54.2%) than enlisted (*n* = 176, 45.8%). The sample was slightly more White than the over-all military racial ethnoracial makeup (Department of Defense, 2020); however, the sample represents military enlistment in the Central US region from which the sample was drawn. The sample was primarily comprised of SMs in the Army (*n* = 347, 90.4%) or Army Reserve (*n* = 15, 3.9%). All participants had a diagnosis of mild traumatic brain injury. Diagnosis of mTBI was operationally defined via the Department of Defense [DoD]/Veteran Affair [VA] criterion, which includes the following after injury: loss of consciousness 30 min or less; post-traumatic amnesia 24 hour or less; self-reported alteration of consciousness/mental state lasting up to 24 hours; or Glasgow Coma Scale score ≥ 13. Brain injury across the sample was remote with an average of 85.8 months since the most significant TBI (SD = 75.4) and 61.6 months since the last TBI (SD = 51.8). Most participants had multiple concussions (M = 5.0, SD = 5.7), with more occurring during military service (M= 3.91, SD = 5.2) than outside of service (M = 1.08, SD = 2.3).

Measures

Personality Assessment Inventory. The PAI (Morey, 2007) is a widely used measure of personality and emotional functioning consisting of 344 items answered on a 4-point Likert-type format. It is well validated within neuropsychological samples (e.g., Busse et al., 2014) and has a burgeoning work with those with mTBI. The PAI contains four primary validity scales (Inconsistency [ICN], Infrequency [INF], Positive Impression Management [PIM], and Negative Impression Management [NIM]), as well as 11 clinical scales (each with subscales), 5 treatment scales, and 2 interpersonal scales.

Cognitive Measures. Available concurrently administered psychological testing evaluating common measures of intellectual and memory is used within this paper as criterions across which cognitive performance is contrasted between identified groups. Administered tests include the: Weschler Adult Intelligence Scale-Fourth Edition (WAIS-IV; Wechsler, 2008a; 2008b), Controlled Word Association Task (see Rodriguez-Aranda & Martinussen, 2006 for a meta-analytic review), Animal Naming Task (see Tombaugh et al., 1999), Trail Making Test - Forms A & B (see Tombaugh, 2004), and the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS; Randolph, 1998). In general, these tests cover a broad range on intellectual and executive functioning. For an over-view of these collateral measures, the reader is referred to the technical manuals and substantial research base of each.

Psychological Testing of Performance Validity. Performance validity was assessed on subsamples that included any of three separate measures, the Test of memory Malingering (TOMM; Tombaugh, 1996), the non-verbal Medical Symptom Validity Test (NV-MSVT; Green, 2008), and the Medical Symptom Validity Test (MSVT; Green, 2004). For the TOMM, Trials 1 and 2 were administered and PVT failure was determined using a score of 44 on Trial 2 while

MSVT/NV-MSVT failure was determined using manual instructions (see Carone, 2009 and Wagner & Howe, 2010 for reviews of the MSVT and NV-MSVT, respectively).

Procedure

This study used an anonymized clinical database that included scale scores of administered tests but did not contain item-level information. As such, no reliability coefficients were available for the scales. After receiving IRB approval from the Madigan Army Medical Center, analyses were conducted using T-scores on the main PAI substantive scales (i.e., Clinical, Treatment, Interpersonal scales) of the PAI. To ensure the independence of predictors, clinical subscales were excluded from initial analysis and were retained for post-hoc testing only. Latent Class Analysis (LCA) were planned to calculate 8 potential class structures (1-Class through 8-Class), exceeding previously observed class counts. If model fit had continued to suggest additional classes were needed, which we did not expect given prior work, we planned to add additional class structures for analysis.

LCA analyses were conducted in Mplus 8.0 with a maximum likelihood estimator with robust standard errors to account for any non-normality in the data. A combination of fit estimators was used during interpretation of classes, including: (1) Akaike Information Criteria (AIC), (b) Bayesian Information Criteria (BIC), (3) sample size adjusted BIC, (4) Entropy, (5) Adjusted Lo-Mendal-Rubin (LMR), and class scale means. This approach allows for the integration of a theory-based interpretation of the observed class structures, which is preferred for LCA analysis (see Nylund et al., 2007). Differences between clusters were interpreted as meaningful only if means were both statistically ($p < .01$) and clinically different (i.e., exceeded a medium effect size or a $T \geq 5$; Rosenthal et al., 2000) *or* if there were statistically distinct rates of elevation at recommended cut scores based on chi-square analysis and a medium size effect difference using Kramer's *V*. We did not use family-wise corrected significance because of our reliance on clinical significance of differences, which are less influenced by sample size and better reflect meaningful distinctions. *A priori* criteria were also used to determine the viability of identified clusters, including that they contain at least 10% of the sample to avoid the final solution from having small and uninterpretable clusters. Following the identification of a reliable class solution, we planned to examine differences across all PAI scales (i.e., both those analyzed, as well as the subscales which comprise each clinical scale). Post-hoc analyses were also planned on cognitive testing (e.g., WAIS-IV, Trail Making Test), performance validity testing (PVT), and positive neuroimaging (i.e., indicating a structural abnormality) frequency amongst those participants with any such data. Positive imaging was determined by the chart coding of the radiologist who evaluated the neuroimaging, which was recorded at the time of evaluation.

Results

LCA models for 2-Class through 8-class solutions were conducted (Table 2). While the three-class solution produced slightly improved fit on some key statistics (e.g., Δ AIC = 590.98, Δ BIC = 515.97, Δ Adjusted BIC = 576.25), it also produced worst classification (entropy). Moreover, when class structures (i.e., mean score patterns) for the various class solutions were evaluated, these models (including the three-class solution) demonstrated a broad, progressive increase in levels of scale scores, rather than discrete pattern(s) of symptoms which are distinct and theoretically interpretable. Similar patterns of a general distress elements were observed across the other LCA models, and changes in fit statistics were small when compared to observed fit values in other class solutions (i.e., 2-Class/3-Class ΔAIC = 590.98 while overall AIC in the 3-class is 49,880.90). These results observed across the multiple class structures identified and analyzed indicate that the underlying symptom pattern evidenced within analysis is likely dimensional rather than categorical. Thus, a 2-class solution produced the most parsimonious model for the same pattern (i.e., symptom severity) observed across all LCA analyses and offers a concrete way to synthesize how specifically pronounced patterns of pathology. differences in scale mean scores for the 2-class structure (Table 3) on the clinical scales (*d* = .09 to 2.54) were large/very large in effect. Most pronounced (i.e., highest T-score and most frequent clinical elevation [≥T70]) amongst scales assessing internalizing and thought disordered pathology. The magnitude of effect differences between classes varied as a function of the number of extracted classes.

Post-hoc analyses (Table 4) were conducted using the final grouping for each participant from the 2-Class classification to determine specific symptom sets that are likely driving pathology elevations on the overall clinical scales used for the LCA. PAI Clinical Subscales, which were not included in the LCA and that are part of overall Clinical scale score, produced more subdued group differences, with approximately half demonstrating medium (35%) to large (10%) range in the 2-class solution. These differences were primarily across the internalizing and thought disordered scales. Similarly, neurocognitive testing found less frequent differences between these pathology-level groups, with some evidence of slightly lower intellectual functioning in the medium to large effect range. However, this pattern was not conclusive and did not emerge across all cognitive testing data. Evidence also suggests minor differences on PVT performance, with one PVT (NV-MSVT) differing significantly across the two-class solution. Table 3 and 4 also contains descriptive information about outcome measures for primary analyses, as well as those conducted post-hoc using the identified class structure.

Discussion

This study investigated the potential emergence of different symptom classes using latent class analysis (LCA) on the Personality Assessment Inventory (PAI) within an active-duty military population meeting DOD/VA requirements for diagnosis of mTBI. Prior literature has found distinct symptom groups in non-military TBI populations across broadband measures; however, the number and nature of these groups have differed across studies (e.g., Goldsworthy & Donders, 2019; Johnstone et al., 2022). Indeed, some researchers have found no need for subgroups at all and recommend unified comparison groups (Childs et al., 2022). Moreover, prior studies have investigated only the main clinical scales of the PAI and have excluded extra-test variables which provide interpretive validity to ensure class structures differ.

We found that a two-class solution (low/high symptom groups) could provide adequate explanation for the data patterns found within our analysis wherein the low symptom group had scores that were largely normative, and the high symptom group had notable and frequent elevations on scales measuring somatic concerns, depression, anxiety, anxiety-related disorders, and paranoia. While the PAI does not having a general distress scale like comparable broadband personality measures, the frequency of elevations seen in the high symptom groups appears marked by general distress (see Morey, 1996's discussion of first factor response patterns on the PAI). This two-class solution is representative of an underlying symptom severity trait, rather than a distinct set of classes which should be interpreted. Thus, we caution researchers about interpreting them as truly distinct because they merely provide parsimoniously fitted model that describes a dimensional pattern (severity), highlighting that neuropsychology clients with remote mTBI demonstrate no distinct subclass patterns. These findings differ from previously observed three- (Demakis et al., 2007; Velikonja et al., 2010) and four-class (Johnstone et al., 2022;

Kennedy et al., 2015) solutions. In contrast, interpretive conclusions are consistent with Childs et al (2022).

Broadly, patterns observed across LCA models (2-class through 7-class) suggest that rather than distinct and meaningful classes, symptom severity levels better explain class structures. In other words, identified classes increase as a function of general symptom severity (defined by elevation frequency and greater mean scores across PAI scores) than distinct patterns, which has previously been interpreted as the causal factor for the underlying class structures. This finding of symptom severity as an explanation for class groups (rather than distinct patterns) aligns with our hypotheses that existing classes would not replicate and with literature on latent class analysis more broadly (e.g., Ingram et al., 2021). In addition to findings related to the classification of individuals into groups, we examined the relationship between symptoms and extra-test performance, as measured by standalone performance validity and cognitive testing. Somewhat consistent with our hypotheses and past research (Jurick et al., 2019), we found that the higher symptomatic individuals would have a greater percentages of PVT failure and lower cognitive testing scores. However, differences were generally modest and were not evident on all tests, leading to this hypothesis to not be fully supported. Our hypothesis that those with lower symptom patterns would have less serious injuries (as defined by abnormal imaging or by concussion severity) was not supported. In general, this study provides three specific implications for the evaluation of active-duty individuals with a history of mild TBI: (1) specific class models observed in prior research are not seen in this active-duty sample, (2) symptom severity explains the emergence of various class structures, and (3) symptoms reported within the high symptom group are broad and include experiences outside of domains expected in those with a history of brain injury.

A major implication of our work is that clinicians evaluating active-duty personnel with mTBI should not necessarily expect distinct class structures found elsewhere in the PAI LCA literature (e.g., Goldsworthy & Donders, 2019; Johnstone et al., 2022). Such a finding is particularly warranted in a remotely injured sample, consistent with probable recovery from brain injury (Iverson, 2005). Indeed, in these mTBI populations with remote injury, comparison groups may be created by collapsing individuals into a single comparison group (see Childs et al., 2022). Accordingly, clinicians should not necessarily base recommendations and treatment needs on similar research in other samples and rely on the scales of the PAI based on standard interpretation, for those with remote injury. We based our interpretation of results on the idea that parsimonious explanations are preferred in data analysis (e.g., that one explanatory factor [severity] is better than multiple and inconsistent patterns [numerous class structures observed across various samples]). Indeed, this conclusion may also apply to non-military samples as well because of over-extraction difficulties in the approaches previously utilized; however, this possibility should be verified in a non-military sample with less remote injury.

Our conclusion of non-distinct classes somewhat contrasts the conclusions found a four-class structure (no distress, moderate distress, high distress, and somatic distress) found by Kennedy et al. (2015). However, outside of the somatic distress class, classes identified by Kennedy and colleagues may also be interpreted as reflecting a symptom severity factor (e.g., low, moderate, high distress) that would emerge as additional classes are added to the model. Kennedy et al's use of cluster analysis, which does not provide a clear set of fit statistics by which to judge meaningful model additions, may have also led to an over-extraction or misinterpretation of identified classes (e.g., Aguerrevere et al., 2018; Soldino et al., 2019). An added difficulty to interpreting less-reliably cluster derived classes is that when studies do not include data on which to evaluate potential class structures (e.g., criterion validity) outside of the scales used to create the classes, interpretation becomes difficult. Class structures on the PAI are often better explained by continuous, latent constructs measuring general mental health/distress (see Ingram et al., 2021). Differences in injury severity may have also impacted findings as 33% of Kennedy's sample had moderate or severe TBI and there were no such cases in our study (although most of those individuals fell in the low distress group as expected [see Jurick et al., 2019]).

Johnstone et al. (2022) found a similarly broad distress related theme in their evaluation of active-duty personnel, containing most items analyzed within their study. The additional groups included scale item content from included internalizing, externalizing, and thought disordered content, despite such content being clinically unique (Kotov et al., 2017). Groups did not include most of the items for the scales from which the group's content theme is measured. For instance, the "Substance Misuse" group had a total of 6 items while the PAI's substance use scales from which items were identified for that group (Alcohol Problems [ALC] and Drug Problems [DRG]) contain substantially more items. Moreover, the sample's mean scores for these scales approximate a T-score of 50 and have standard deviations comparable to the PAI normative sample. Thus, it is likely that their results rely on over-extraction of item content rather than distinct and meaningful symptom groups. The fact that they used exploratory factor analysis limits the interpretative comparability.

Further supporting a general pathology level as the core explanation for distinct class structures at this stage post-injury is the observation that the most pronounced scale score elevations (e.g., ARD [Anxiety Related Disorders] M=69.4; DEP [Depression] M=72.6; SCZ [Schizophrenia] M=68.9) fell outside expected symptoms associated most with brain injury (i.e., cognitive and somatic symptoms). It is possible that these affective features are the result of brain injury given that similar internalizing distress problems are frequently observed in this population (see Whiteside et al., 2012); however, the subscales driving those elevations do not fully support that conclusion. Instead, elevations on scales measuring these constructs may reflect issues which may surround that injury (e.g., psychological trauma experiences; Sellbom et al., 2012). For instance, the ARD-T (Anxiety-Related Disorders- Traumatic Stress) scale, which is most associated with traumatic experiences (Bellet et al., 2018), is the ARD subscale that was most elevated. This elevation aligns with past findings that mTBI does not predict PAI symptom elevations. Rather PTSD and trauma experiences may serve as the main predictor for PAI symptom elevation in those with mTBI (Miskey et al., 2015). Likewise, DEP-P (Depression - Physiological), rather than DEP-C (Depression - Cognitive) which appears to underlie the DEP elevation. Conversely, elevations on SCZ are driven primarily by the SCZ-Thought Disorder (SCZ-T) scale, which may reflect cognitive problems common in brain injury, despite the remote injury. Those elements of poorer memory, concentration, and focus are, however, also common amongst a variety of distress-based disorders. While somatization elevations also increased (as did all scales) across different class solutions, it is possible that somatic symptom elevations not being the most pronounced may occur as a result the mTBI recovery process (see Iverson, 2005). However, and more likely given past work with latent class analysis on the PAI (Ingram et al., 2021), is that brain injury doesn't itself produce distinct classes in broadband assessment profiles (e.g., Greiffenstein and Baker, 2001), at least at this stage of recovery (Childs et al., 2022; Iverson, 2005). Rather, occurrence of general psychopathology within the sample and population/setting specific base rates may better predict the types of symptoms which become most evident in the latent groups within these individuals. Even if the underlying cause is attributed to affective and cognitive changes caused by mTBI, these changes still appear as a function of a dimensional pattern of symptom severity rather than distinct class patterns found previously.

Limitations and Future Directions

No study is without limitations. Participants primarily had remote brain injury histories and those injuries were classified as mild in nature. While mild traumatic brain injury is highly prevalent in military samples (e.g., DVBIC, 2016; Frost et al., 2013; Schneider et al., 2018), this study should not be generalized to acute, recent, or more severe cases of brain injury until supported by replication. Accordingly, our results suggesting a latent severity factor differs from prior work on mTBI using the PAI and may reflect patterns only observed in those with mild TBI, despite having similar findings to other PAI latent class research (Ingram et al., 2021). An additional limitation is that this sample is composed of primarily White males and may not generalize to more diverse populations. Additionally, our sample had an equal representation of officers and enlisted, and we did not examine how officer or enlisted status (or the associated service experiences resulting in mTBI) may differ. Thus, further work should determine if rank and associated service experiences (service duties, service length, etc) produce distinct patterns, or impact observed patterns, between groups Future research should continue to incorporate extra-test measures of performance validity and cognitive testing and should evaluate different TBI severities as an impactful factor of class structures.

Moreover, since existing research with those who have a TBI on the PAI is limited and contradictory in both the number and nature of observed classes (see Demakis et al., 2007; Kennedy et al., 2015; Johnstone et al., 2022 Velikonja et al., 2010); research remains needed to determine if differences can be explained solely by the analytic techniques used or if there are underlying sample differences (e.g., TBI severity) that better explain the observed patterns. It is also possible that differences in AD and civilian populations drove some differences in identified class structure. Given prior work highlighting similar difficulties in extracting theoretically meaningful groups using LCA (see Ingram et al., 2021), we believe our interpretation is likely the most robust; however, replication and extension is needed. It is also important to control for invalid responding and not doing so can bias subsequent analytic interpretation (Shura et al., 2021; Hawes, & Boccaccini, 2009).

We controlled for over-reporting by excluding participants exceeding recommended cut-scores (e.g.,, NIM ≥ 110; Morey et al., 2007) as well as those undergoing Medical Board and Temporary Disability evaluations. Such an exclusion approach is not inclusive of all potential metrics to identify invalid responding; however, it provides a middle road between over-exclusion which might not generalize to prior studies and non-exclusion, which biases data analysis. While these exclusions reduced the amount of invalid responding, the high symptom group had a higher percentage of participants failing performance validity tests (32.2% MSVT and 21.1% NVMSVT). Higher invalidity amongst active-duty populations with mTBI (Jurick et 2019) and those with higher symptom experiences is not uncommon or unexpected; however, external motivations can't be ruled out. Subsequent research should evaluate if clustered patterns of SVT/PVT responses product distinct classes. While such an investigation of validity scales is outside the scope of this paper, it would be wise to approach such an undertaking with explicit focus on the newly developed cognitive over-reporting measures for the PAI (Armistead-Jehle & Ingram, in press; Boress et al., 2021; Ingram et al., in press; Gaasedelen et al., 2019; Shura et al., in press). Likewise, using these cognitively focused scales as the sole exclusion criteria prior to LCA provides another direction worthy of independent investigation. Such work would vary from this and prior LCA work in mTBI populations because of the distinct exclusionary approach (see Sweet et al., 2021). These limitations notwithstanding, this study offers adds to the literature on groups of respondents with mTBI on the PAI and offers a potential explanation which allows past, contradictory findings to be explained parsimoniously in those with remote injury.

References

Aguerrevere, L. E., Calamia, M. R., Greve, K. W., Bianchini, K. J., Curtis, K. L., & Ramirez, V. (2018). Clusters of financially incentivized chronic pain patients using the Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2-RF). *Psychological Assessment,* 30(5), 634–644. https://doi.org/10.1037/pas0000509

Arbisi, P. A., Polusny, M. A., Erbes, C. R., Thuras, P., & Reddy, M. K. (2011). The Minnesota Multiphasic Personality Inventory–2 Restructured Form in National Guard soldiers screening positive for posttraumatic stress disorder and mild traumatic brain injury. *Psychological Assessment,* 23(1), 203.

Armistead-Jehle, P., & Ingram, P.B. (In Press). Validating the Memory Complaints Inventory (MCI) detection of improbable cognitive symptoms: establishing preliminary cut scores in a Military sample. *Archives in Clinical Neuropsychology*

Armistead-Jehle, P., Ingram, P. B., & Morris, N. M. (2020). Personality assessment inventory cognitive bias scale: Validation in a military sample. *Archives of Clinical Neuropsychology*, 35(7), 1154-1161.

Bass, C. R., Panzer, M. B., Rafaels, K. A., Wood, G., Shridharani, J., & Capehart, B. (2012). Brain injuries from blast. *Annals of biomedical engineering*, 40(1), 185-202.

Bellet, B. W., McDevitt-Murphy, M. E., Thomas, D. H., & Luciano, M. T. (2018). The utility of the personality assessment inventory in the assessment of posttraumatic stress disorder in OEF/OIF/OND veterans. *Assessment*, 25(8), 1074-1083.

Ben-Porath, Y. S., & Tellegen, A. (2008/11). MMPI–2–RF (Minnesota Multiphasic Personality Inventory–2 Restructured Form): Manual for administration, scoring, and interpretation. Minneapolis, MN: University of Minnesota Press

Bolinger, E., Reese, C., Suhr, J., & Larrabee, G. J. (2014). Susceptibility of the MMPI-2-RF neurological complaints and cognitive complaints scales to over-reporting in simulated head injury. Archives of clinical neuropsychology, 29(1), 7-15.

Boress, K., Gaasedelen, O. J., Croghan, A., Johnson, M. K., Caraher, K., Basso, M. R., & Whiteside, D. M. (2021). Replication and cross-validation of the personality assessment inventory (PAI) cognitive bias scale (CBS) in a mixed clinical sample. The Clinical Neuropsychologist, 1-19.

Busse, M., Whiteside, D., Waters, D., Hellings, J., & Ji, P. (2014). Exploring the reliability and component structure of the personality assessment inventory in a neuropsychological sample. The Clinical Neuropsychologist, 28(2), 237-251.

Carone, D. A. (2009). Test review of the medical symptoms validity test. Applied Neuropsychology, 16(4), 309–311. https://doi.org/10.1080/09084280903297883

Childs, A., Bertisch, H., Talis, E., Ricker, J. H., & Rath, J. F. (2022). Development of an MMPI reference group for outpatients with persisting symptoms following mild TBI. Brain Injury, 1-7.

Comper, P., Bisschop, S. M., Carnide, N., & Tricco, A. (2005). A systematic review of treatments for mild traumatic brain injury. Brain injury, 19(11), 863-880.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, 52(4), 281.

Defense and Veteran Brain Injury Center Annual Report. (2019). Annual Report 2019. Silver Spring, MA: Defense and Veterans Brain Injury Center.

Denning, J. H., & Shura, R. D. (2019). Cost of malingering mild traumatic brain injury-related cognitive deficits during compensation and pension evaluations in the veterans benefits administration. *Applied Neuropsychology: Adult*, 26(1), 1-16.

Department of Defense (2020). 2020 Demographics: Profile of the Military Community. Retrieved from https://download.militaryonesource.mil/12038/MOS/Reports/2020-demographics-report.pdf

Frost, R. B., Farrer, T. J., Primosch, M., & Hedges, D. W. (2013). Prevalence of traumatic brain injury in the general adult population: a meta-analysis. *Neuroepidemiology*, 40(3), 154-159.

Gaasedelen, O.J., Whiteside, D.M., Altmaier, E., Welch, C., & Basso, M.R., (2019). The construction and initial validation of the Cognitive Bias Scale for the Personality Assessment Inventory. The Clinical Neuropsychologist

Gervais, R. O., Wygant, D. B., Sellbom, M., & Ben-Porath, Y. S. (2011). Associations between symptom validity test failure and scores on the MMPI–2–RF validity and substantive scales. *Journal of Personality Assessment, 93*(5), 508-517.

Goldsworthy, R., & Donders, J. (2019). MMPI-2-RF patterns after traumatic brain injury. *Psychological Assessment*, 31(9), 1145.

Green, P. (2004). Green's Medical Symptom Validity Test (MSVT) for Microsoft Windows: User's manual. Edmonton, Canada: Green's Publishing.

Green, P. (2008). Manual for the Nonverbal Medical Symptom Validity Test. Edmonton, Canada: Green's Publishing

Hawes, S. W., & Boccaccini, M. T. (2009). Detection of overreporting of psychopathology on the Personality Assessment Inventory: a meta-analytic review. *Psychological Assessment*, 21(1), 112.

Hoge, C. W., McGurk, D., Thomas, J. L., Cox, A. L., Engel, C. C., & Castro, C. A. (2008). Mild traumatic brain injury in US soldiers returning from Iraq. *New England journal of medicine*, 358(5), 453-463.

Ingram, P.B., Armistead-Jehle, P., Herring, T.T., & Morris, N.M., (In Press). Cross Validation of the Personality Assessment Inventory (PAI) Cognitive Bias Scale of Scale (CBS-SOS) over-Reporting Indicators in a Military Sample. *Military Psychology*

Ingram, P. B., Armistead-Jehle, P., Morris, N. M., & Herring, T. (2022). Cross Validation of the Personality Assessment Inventory (PAI) Cognitive Bias Scale of Scale (CBS-SOS) over-Reporting Indicators in a Military Sample. *Archives of Clinical Neuropsychology*. Online First Publication

Ingram, P.B., Golden, B.L., & Armistead-Jehle, P. (2020). Evaluating the Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF) over-reporting scales in a military neuropsychology clinic. *Journal of Experimental and Clinical Neuropsychology,* 42(3), 263-273. http://doi.org/10.1080/13803395.2019.1708271

Ingram, P.B., Mosier, N.J., Sharpnack, J.D., & Golden, B. (2021). Evaluating symptom endorsement typographies of trauma-exposed veterans on the Personality Assessment Inventory: A latent profile analysis. *Current Psychology,* 40(11), 5267-5277. https://doi.org/10.1007/s12144-019-00486-5

Ivins, B. J., Schwab, K. A., Warden, D., Harvey, L. S., Hoilien, M. M., Powell, C. J., ... & Salazar, A. M. (2003). Traumatic brain injury in US Army paratroopers: prevalence and character. *Journal of Trauma and Acute Care Surgery*, 55(4), 617-621.

Jones, A., Ingram, M. V., & Ben-Porath, Y. S. (2012). Scores on the MMPI-2-RF scales as a function of increasing levels of failure on cognitive symptom validity tests in a military sample. *The Clinical Neuropsychologist,* 26(5), 790-815.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. Biochemia medica, 22(3), 276-282.

Meyers, J. E., Miller, R. M., Haws, N. A., Murphy-Tafiti, J. L., Curtis, T. D., Rupp, Z. W., ... & Thompson, L. M. (2014). An Adaptation of the MMPI-2 Meyers Index for the MMPI-2-RF. *Applied Neuropsychology: Adult*, 21(2), 148-154.Morris, N. M., Ingram, P. B., & Armistead-Jehle, P. (2022). Relationship of personality assessment inventory (PAI) over-reporting scales to performance validity testing in a military neuropsychological sample. *Military Psychology*, 1-10.

F, L. P. (1950). The logical and mathematical foundation of latent structure analysis. *Studies in Social Psychology in World War II Vol. IV : Measurement and Prediction*, 362–412.

Miskey, H. M., Shura, R. D., Yoash-Gantz, R. E., & Rowland, J. A. (2015). Personality Assessment Inventory profiles of veterans: Differential effects of mild traumatic brain injury and psychopathology. *Brain Imaging and Behavior*, *9*(3), 461–471. https://doi.org/10.1007/s11682-015-9391-7

Randolph, C. (1998). Repeatable Battery for the Assessment of Neuropsychological Status (RBANS). San Antonio, TX: Psychological Corporation.

Rodríguez-Aranda, C., & Martinussen, M. (2006). Age-related differences in performance of phonemic verbal fluency measured by Controlled Oral Word Association Task (COWAT): a meta-analytic study. Developmental neuropsychology, 30(2), 697-717.Rona, R. J., Jones, M., Fear, N. T., Hull, L., Murphy, D., Machell, L., ... & Wessely, S. (2012).

Mild traumatic brain injury in UK military personnel returning from Afghanistan and Iraq: cohort and cross-sectional analyses. *The Journal of head trauma rehabilitation*, 27(1), 33-44.

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). Contrasts and effect sizes in behavioral research: A correlational approach. Cambridge University Press.

Sarstedt, M., & Mooi, E. (2014). Cluster analysis. In M. Sarstedt, & E. Mooi (Eds.), A concise guide to market research: The process, data, and methods using IBM SPSS statistics (pp. 273–324). Berlin, Germany: Springer Verlag. http://dx.doi.org/10.1007/978-3-642-53965-7_9

Schneider, A. L., Wang, D., Ling, G., Gottesman, R. F., & Selvin, E. (2018). Prevalence of self-reported head injury in the United States. *The New England journal of medicine*, 379(12), 1176.

Schroeder, R. W., Baade, L. E., Peck, C. P., VonDran, E. J., Brockman, C. J., Webster, B. K., & Heinrichs, R. J. (2012). Validation of MMPI-2-RF validity scales in criterion group neuropsychological samples. *The Clinical Neuropsychologist,* 26(1), 129-146.

Sellbom, M., Lee, T. T., Ben-Porath, Y. S., Arbisi, P. A., & Gervais, R. O. (2012). Differentiating PTSD symptomatology with the MMPI-2-RF (Restructured Form) in a forensic disability sample. *Psychiatry Research*, 197(1-2), 172-179.

Shura, R. D., Denning, J. H., Armistead-Jehle, P., & Axelrod, B. N. (2021). Validity testing within the VA and active duty settings. In R. W. Schroeder & P. K. Martin. (Eds.), Neuropsychological Validity Testing in Clinical Settings: A Handbook for Non-Forensic Practice. Guilford Press.

Shura, R., Ingram, P.B., Miskey, H.M., Martindale, S.L., Rowland, J.A., & Armistead-Jehle, P.

(In Press). Validation of the Personality Assessment Inventory (PAI) Cognitive Bias (CBS) and Cognitive Bias Scale of Scales (CB-SOS) in a Post-Deployment Veteran Sample. *The Clinical Neuropsychologist*

Soldino, V., Carbonell-Vayá, E. J., Perkins, D., & Tibau, X. A. (2019). MCMI-III profiles of convicted contact sexual offenders: A cluster analysis. *Journal of forensic psychology research and practice,* 19(5), 393-408.

Tarescavage, A. M., Wygant, D. B., Gervais, R. O., & Ben-Porath, Y. S. (2013). Association between the MMPI-2 Restructured Form (MMPI-2-RF) and malingered neurocognitive dysfunction among non-head injury disability claimants. *The Clinical Neuropsychologist*, 27(2), 313-335.

Tellegen, A., Ben-Porath, Y. S., McNulty, J. L., Arbisi, P. A., Graham, J. R., & Kaemmer, B. (2003). The MMPI-2 Restructured Clinical (RC) Scales: Development, validation, and interpretation. Minneapolis, MN: University of Minnesota Press

Terrio, H., Brenner, L. A., Ivins, B. J., Cho, J. M., Helmick, K., Schwab, K., ... & Warden, D. (2009). Traumatic brain injury screening: preliminary findings in a US Army Brigade Combat Team. *The Journal of head trauma rehabilitation*, 24(1), 14-23.

Thomas, M. L., & Youngjohn, J. R. (2009). Let's not get hysterical: Comparing the MMPI-2 validity, clinical, and RC scales in TBI litigants tested for effort. *The Clinical Neuropsychologist*, 23(6), 1067-1084.

Tombaugh, T. N. (1996). Test of memory malingering.

Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. Archives of clinical neuropsychology, 14(2), 167-177.

Tombaugh, T. N. (2004). Trail Making Test A and B: normative data stratified by age and education. Archives of clinical neuropsychology, 19(2), 203-214.

Tylicki, J. L., Tarescavage, A. M., & Wygant, D. B. (2021). Assessment of malingering among head injury litigants with the MMPI-2, MMPI-2-RF, and MMPI-3. In Detection of Malingering during Head Injury Litigation (pp. 289-307). Springer, Cham.

Wager, J.G., & Howe, L.L.S. (2010). Nonverbal medical symptom validity test: Try faking now! Applied Neuropsychology, 17, 305-309.

Wechsler, D. (2008a). Wechsler Adult Intelligence Scale—Fourth Edition. San Antonio, TX: Pearson Assessment.

Wechsler, D. (2008b). Wechsler Adult Intelligence Scale—Fourth Edition: Technical and interpretive manual. San Antonio, TX: Pearson Assessment

Whiteside, D. M., Galbreath, J., Brown, M., & Turnbull, J. (2012). Differential response patterns on the Personality Assessment Inventory (PAI) in compensation-seeking and non-compensation-seeking mild traumatic brain injury patients. Journal of Clinical and Experimental Neuropsychology, 34(2), 172-182.

Whitman, M. R., Tylicki, J. L., Mascioli, R., Pickle, J., & Ben-Porath, Y. S. (2021). Psychometric properties of the Minnesota Multiphasic Personality Inventory-3 (MMPI-3) in a clinical neuropsychology setting. *Psychological Assessment*, 33(2), 142.

Windgassen, S., Moss-Morris, R., Goldsmith, K., & Chalder, T. (2018). The importance of cluster analysis for enhancing clinical practice: an example from irritable bowel syndrome. Journal of Mental Health, 27(2), 94-96.

Wojcik, B. E., Stein, C. R., Bagg, K., Humphrey, R. J., & Orosco, J. (2010). Traumatic brain injury hospitalizations of US army soldiers deployed to Afghanistan and Iraq. *American journal of preventive medicine*, 38(1), S108-S116.

Worthen, M. D., & Moering, R. G. (2011). A practical guide to conducting VA compensation and pension exams for PTSD and other mental disorders. *Psychological Injury and Law,* 4(3), 187-216.

Young, G. (2020). Thirty complexities and controversies in mild traumatic brain injury and persistent post-concussion syndrome: A roadmap for research and practice. Psychological Injury and Law, 13(4), 427-451.

Youngjohn, J. R., Wershba, R., Stevenson, M., Sturgeon, J., & Thomas, M. L. (2011). Independent validation of the MMPI-2-RF somatic/cognitive and validity scales in TBI litigants tested for effort. *The Clinical Neuropsychologist,* 25(3), 463-476.