Another Perspective of the Miller Forensic Assessment of Symptoms Test – Part I: A Narrative Review

D. Detullio

ABSTRACT

Reference [1] presented a skewed perspective of the M-FAST literature base and provided the flawed conclusion that the M-FAST should no longer be used in practice. In an attempt to correct the many issues with [1], this article provides a narrative review of the strengths and weaknesses of research findings for the M-FAST interpretation as well as reviews methodological concepts underlying feigning research.

The M-FAST was designed to screen for potential feigning of psychiatric symptoms. It was not designed to conclude that an examinee is feigning or malingering psychiatric symptoms. A positive result on the M-FAST only indicates that additional data needs to be collected to make the aforementioned conclusions. Applying the M-FAST in any other way is a serious error on the part of the user.

The research literature thus far generally supports the use of the M-FAST cut-off as a screening measure for possible feigning of psychiatric symptoms. However, there are scenarios when the M-FAST may not operate as efficiently, and these scenarios are discussed. Reference [1] misrepresented the purpose of the M-FAST as well as research findings on the M-FAST. Therefore, [1] should be read with great caution.

Keywords: Feigning, Malingering, Miller Forensic Assessment of Symptoms Test (M-FAST).

I. INTRODUCTION

Reference [1] inaccurately concluded that the M-FAST should no longer be used in practice. Many of [1]’s criticisms regarding the M-FAST can be negated with a review of the professional manual [2], especially criticisms regarding the purpose, administration, interpretation, and development of the M-FAST. Some of [1]’s criticisms could also apply to nearly any standardized psychological instrument, which makes such criticisms useless in their generalities. To fully address the problems in [1], two articles were written: the current article narratively reviews the M-FAST literature; the companion article [55] will provide an updated quantitative analysis to correct the problems with [1] and limitations of [3]. Even with two articles, though, the multitude of issues in [1] could not be fully addressed. Therefore, readers are recommended to review additional sources regarding the theories of feigning and methodologies for studying feigning (e.g., [4], [5]).

Perhaps, the most important piece of information to remember about the M-FAST is that it is a screening instrument to detect potential feigning of psychiatric symptoms. It should never be used to conclude that an examinee is feigning or malingering mental illness. For a discussion regarding the differences between feigning and malingering as well as the role of psychological instruments, see [4]. The M-FAST was also not designed for detecting potential feigning of cognitive or medical conditions. So, it would be inappropriate to use it for these purposes.

II. THE VALIDITY OF THE M-FAST

Reference [1] presented a skewed perspective of the research base on the M-FAST. Reference [1] favored data that supported the supposed lack of validity for the M-FAST while attempting to negate data that supported the validity of the M-FAST. The following subsections will provide a narrative review of study results on the M-FAST as well as methodological concepts. The study results will be reviewed according to the type of feigning design: simulation, partial criterion design (PCD), and known-groups comparison (KGC). For detailed descriptions of these designs, review [4], [5].

A. M-FAST Simulation Designs

There are several studies on the M-FAST that used a simulation methodology, which was the primary type of study included within [1]. As [1] correctly recognized, simulation designs which do not incorporate a clinical comparison group provide limited evidence for the validity of a feigning instrument. Therefore, simulation designs without a clinical comparison group will not be discussed within this section, but such designs could be considered within the context of results from other simulation designs (see below). Simulation studies with clinical comparison groups have generally supported the validity of the M-FAST total score as a

Submitted: November 13, 2021
Published: December 21, 2021
ISSN: 2593-8339
DOI: 10.24018/ejmed.2021.3.6.1142
D. Detullio
Western State Hospital, Washington, United States
(e-mail: detullio.david@gmail.com)*Corresponding Author

DOI: http://dx.doi.org/10.24018/ejmed.2021.3.6.1142
Vol 3 | Issue 6 | December 2021
The previous paragraphs focused on discussing the general findings of simulation studies with an emphasis on the comparison group that was utilized. However, the procedures for instructing participants to feign in simulation designs are important elements as well because such procedures can improve the interpretability and generalizability (i.e., external validity) of the results. Reference [5] provided a checklist of factors that could improve the rigor of the manipulation (i.e., feigning instructions) in simulation designs. A subset of these factors was used to evaluate the feigning methods of simulation studies with a clinical comparison group as well as those simulation designs without a clinical comparison.

Reference [1] criticized many studies as not having a comparison group of “very severely ill psychiatric patients” (p. 89). It is correct that many of the studies [1] focused on did not have such a comparison group. This is because [1] decided to exclude multiple studies that did have such a comparison group (e.g., [20]). Based on the results of [11] and [15], [1] indicated “the diagnostic probability of M-FAST misclassifying such genuine patients with severe mental illness as malingers is about 50%” and concluded “the data indicate that the M-FAST should not be used on such patient groups at all” (p. 89). Reference [1]’s conclusion from these two studies is flawed considering the M-FAST is not intended for making a final determination as to malingering because it is a screening instrument, and the operating characteristics of the cut-off for the M-FAST total score appeared sufficient for a screening instrument. Moreover, [1]’s specificity estimates for the two aforementioned studies are underestimates. Lastly, [15] used a translated version of the M-FAST that was administered to participants in Armenia. These cultural factors could have impacted the specificity estimate similar to [18], which negates [1]’s implication that the results were due entirely to genuine psychopathology.

TABLE I: M-FAST SIMULATION STUDIES

<table>
<thead>
<tr>
<th>Ref</th>
<th>Sens</th>
<th>Spec</th>
<th>Comp</th>
<th>Specific Real-World Scenario</th>
<th>Familiar with Simulation Context</th>
<th>Caution</th>
<th>Coaching</th>
<th>Internal Motivation</th>
<th>Incentives for feigning</th>
<th>Manipulation Check</th>
<th>Relevance of Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>[6]</td>
<td>0.51</td>
<td>0.87</td>
<td>NC</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>[7]</td>
<td>0.83</td>
<td>0.97</td>
<td>NC</td>
<td>Yes</td>
<td>Yes</td>
<td>DC</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>[8]</td>
<td>0.91</td>
<td>1.00</td>
<td>NC</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>[9]</td>
<td>0.67</td>
<td>0.95</td>
<td>NC</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Positive Incentive</td>
<td>Positive Incentive</td>
</tr>
<tr>
<td>[10]</td>
<td>0.78</td>
<td>0.79</td>
<td>C</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>[2]</td>
<td>0.93</td>
<td>1.00</td>
<td>NC</td>
<td>No</td>
<td>No</td>
<td>DC</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>[11]</td>
<td>0.90</td>
<td>0.69</td>
<td>C</td>
<td>Yes</td>
<td>No</td>
<td>BC</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>[12]</td>
<td>0.85</td>
<td>1.00</td>
<td>NC</td>
<td>No</td>
<td>No</td>
<td>BC</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>[13]</td>
<td>0.90</td>
<td>1.00</td>
<td>NC</td>
<td>Yes</td>
<td>No</td>
<td>DC</td>
<td>Combination</td>
<td>No</td>
<td>Positive Incentive</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>[14]</td>
<td>0.86</td>
<td>0.95</td>
<td>NC</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>[15]</td>
<td>1.00</td>
<td>0.57</td>
<td>C</td>
<td>Yes</td>
<td>No</td>
<td>DC</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>[16]</td>
<td>0.86</td>
<td>0.68</td>
<td>NC</td>
<td>Yes</td>
<td>No</td>
<td>DC</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>[17]</td>
<td>0.82</td>
<td>0.73</td>
<td>NC</td>
<td>Yes</td>
<td>Yes</td>
<td>DC</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>[18]</td>
<td>0.86</td>
<td>0.37</td>
<td>C</td>
<td>Yes</td>
<td>Yes</td>
<td>DC</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Note. Sensitivity and specificity values in italics were estimated using the method from [1]. The studies were coded using factors from [5] (pp. 594 - 601). M-FAST = Miller Forensic Assessment of Symptoms Test; Ref = reference; Sens = sensitivity; Spec = specificity; Comp = Comparison Group; NC = nonclinical; C = clinical; DC = detection-caution; BC = believable caution.

DOI: http://dx.doi.org/10.24018/ejmed.2021.3.6.1142
group (see Table I and [5]). Focusing on the former simulation designs, all the feigning instructions provided a specific scenario that could occur within real-world settings (e.g., feigning to not be held responsible for a criminal charge), and the majority of instructions cautioned to present as believable and/or cautioned that exaggerating could be detected. Such cautious could increase the likelihood that participants would pay greater attention to their simulated presentation [5]. Reference [10] provided participants with PTSD symptoms to memorize as a means to increase their chances of successfully feigning (i.e., symptom-coaching). However, as discussed in [5], detection-based coaching is typically more effective than symptom-based coaching, and therefore, the sensitivity observed in [10] may be greater than what would have been observed if detection-based coaching was used. Two simulation studies with a clinical comparison group had manipulation checks, and two studies with a clinical comparison group provided a positive incentive for those who successfully feigned symptoms. Manipulation checks are important to simulation designs because they ensure that participants appropriately followed the feigning instructions. If manipulation checks are not completed, it is unknown if participants followed the instructions they were provided, and it is possible that they did not feign as instructed. This would reduce the internal validity for simulation designs. Incentives for successfully feigning are important considerations because incentives could increase participants’ effort to feign, which would enhance external validity given the motivation to feign in real-life settings. Finally, only one study [18] provided a scenario that would be familiar to the feigning participants. None of the individual simulation studies with a clinical comparison group had all the selected factors from [5] that would increase the generalizability of the sensitivity values, but the strengths of some studies counter the weaknesses of others.

As previously discussed, [5] considered simulation designs without a clinical comparison group to be flawed because it would not be known if the differences between groups is due to feigning or psychopathology. However, there are simulation designs which omitted clinical comparison groups that are still worthwhile to consider because they possessed several factors which might increase the generalizability of the sensitivity estimates. Sensitivity values also do not require a comparison group to calculate, which reduces the limitation of lacking a clinical comparison group. Reference [9] included simulation and KGC components within their study. Focusing on only the simulation component of [9], there was no clinical comparison group. As shown in Table I, the methodology for the simulated feigning condition in [9] was quite strong in that the only element it did not have was some form of coaching. In comparison to the sensitivity for other studies, the estimated sensitivity from [9] was at the lower end, and lower than what one may hope for in a screening instrument. Although, the sensitivity value needed to be estimated using the method from [1], which may have impacted said value. Reference [9] did report a sensitivity of 0.74 for the M-FAST total score at the recommended cut-off ([2], p. 12). It was a bit unclear though if this sensitivity value incorporated those participants who were part of the KGC design, and since this discussion focused purely on simulation designs, it was decided to estimate the sensitivity value for the simulation group. It is also unknown whether the simulators were successful in their feigned presentation (see below). References [12] and [13] are relevant in the context of the sensitivity for the M-FAST cut-off because both studies included coaching in the study design. [12] used detection-based coaching, and [13] used a combination of detection-based as well as symptom-based coaching. However, the coaching in [12] appeared somewhat limited. This raises the question of whether the sensitivity for the M-FAST cut-off would have been lower with more extensive detection-based coaching. Reference [13] had many elements to increase the external validity of the simulation component, and if the feigning group was familiar with the context of the feigning scenario, it would have been similar in strength to [9]. The scenario presented to simulators in [12] did not appear to be particularly specific. Consequently, it is unknown if the participants had any familiarity with the presented scenario. Given that the participants in [12] were prisoners, a scenario involving the legal system could have provided greater familiarity, which was the case for [7] and [17]. However, [7] and [17] lacked incentives or manipulation checks. The simulation design with the lowest sensitivity was [6], which instructed simulators to present as though they were experiencing “emotional difficulties” (i.e., PTSD symptoms) subsequent to a motor vehicle accident (p. 244). This is not a surprising finding given [2] discussed how the M-FAST total score may perform poorly in detecting specific symptomatology. When simulators are compared to a group without pathology on a feigning measure, the differences could be due to feigning or genuine pathology [5]. By focusing on only sensitivity values for the simulator groups, as opposed to comparisons between groups, these results can be informative. Nonetheless, the results should still be considered in the context of simulation studies which incorporated a clinical comparison group.

An aspect of simulation designs that tends to be overlooked is whether simulators were successful in presenting themselves as mentally ill. That is, did simulators avoid detection on the feigning measure(s) and still demonstrated a sufficient number of symptoms required for psychiatric diagnoses? In the real-world, feigners need to endorse enough genuine symptoms while avoiding questions intended to detect feigning to achieve their goals (e.g., avoid prosecution). Therefore, a measure of genuine symptoms could be incorporated in simulation designs to identify if simulators report enough genuine symptoms to present as mentally ill while avoiding questions to detect feigning. Otherwise, simulators may simply avoid endorsing the majority of questions, whether they are genuine symptoms or not. None of the studies reviewed here sought to identify if simulators were successful in feigning with respect to the M-FAST alone. Looking to research on other measures (e.g., [21]) it has been observed that simulators who are not detected by feigning measures do not always endorse sufficient levels of clinical symptoms, but this may depend on the symptoms. If measures of genuine symptoms were incorporated in the studies reviewed above, it could have revealed whether those who were not detected as potentially feigning by the M-FAST still reported enough symptomatology to present themselves as experiencing psychopathology. Future studies on the M-FAST might
consider incorporating measures of genuine symptomatology to glean such information.

Simulation designs are meant to increase the ability to make causal attributions regarding the scores on feigning measures. The results from the reviewed studies generally support the validity of the M-FAST total score as a screening instrument for potential feigning. All the simulation designs could be criticized for lacking in generalizability, with some more vulnerable to this criticism than others (see Table I). However, the strengths in some studies can counter the weaknesses in others. Nonetheless, the results from simulation studies should be interpreted in the context of KGCs and PCDs because such designs have improved external validity. It was interesting that the internal motivation factor was not present in any simulation design, but it is possible that this was an artifact of the coding. That is, a limited number of examples from [5] were relied upon for coding this factor (e.g., “Can you beat the test?,” p. 598), and it is possible that such examples did not provide the full range of internally motivating instructions.

The above studies could have been coded differently using the factors from [5] in that stricter, more well-defined operationalizations of the factors could have been applied or broader, more general operationalizations could have been developed. Moreover, some factors for studies could have been coded in either direction and required a judgment call (e.g., [6], [11], [10]). This shows the subjectivity in coding studies for narrative and meta-analytic reviews [22]. Nonetheless, it is believed that the coding decisions have some logical support [5], [23].

B. A Prelude to M-FAST KGCs and PCDs

Many KGCs of the M-FAST included the Structured Interview of Reported Symptoms (SIRS). Reference [1] excluded all of such studies with poor justifications. Reference [24] performed a meta-analysis of the original SIRS [25], and found poorer operating characteristic for the SIRS classifications when compared to initial validation studies. Reference [1] used the results from this meta-analysis, in part, to disqualify studies on the M-FAST which used the SIRS to identify feigning groups. As discussed in [24], the lowered specificity could be indicative of “shrinkage” that can occur with cut-scores (p. 104). However, the results from [24] could have been due to the limitations of the studies used in the meta-analysis (see [26]). Therefore, a conclusion regarding the operating characteristics of the SIRS should not be based on the results of [24] alone. As is the case with primary studies, “the perfect research synthesis does not and never will exist” ([27], p. 518). Consumers of research should view meta-analyses as another methodological tool to help obtain a greater understanding of phenomena, but meta-analyses often cannot provide final conclusions to research questions. It could be argued that it was still worthwhile for [24] to include potentially problematic studies because, if similar operating characteristics were observed across studies, then it would support the generalizability of the SIRS across methodological characteristics. Since differences were observed across the studies, the methodological differences between studies could provide hypotheses for why the SIRS did not perform as well on average, but the results should not be used as conclusory evidence for or against the SIRS.

Reference [1] quoted from [28] how “The M-FAST was selected because it is essentially a short form of the SIRS” (p. 933). The M-FAST and SIRS are both structured interviews and have items with similar content as well as structure. Reference [1] used this as additional justification for eliminating studies using the SIRS. It is true that measures with similar methods of measurement have higher correlations [29]. This does not mean that M-FAST studies which used the SIRS should be completely disregarded. Despite the criticisms of the SIRS/SIRS-2 ([24], [30]), at the time of this writing, there has been no other measure as extensively developed to assess for feigning mental illness. Consequently, in the context of KGCs and PCDs, the SIRS would provide the most reliable and valid means of comparing potential feigners and honest responders on the M-FAST total score. Nonetheless, should readers still recognize the possibility for the M-FAST to have greater sensitivity/specificity estimates when the SIRS is used to identify groups? Absolutely, and readers should consider sensitivity/specificity estimates found when using other instruments to identify feigning and comparison groups. Such research will be discussed below, including the weaknesses.

Studies of the M-FAST where the SIRS was the means for identifying groups to compare on the M-FAST total score have been considered to demonstrate the “criterion validity” for the M-FAST (e.g., [2], p. 26). Reference [1] took issue with this designation and noted that the appropriate criterion for the M-FAST would be “the capacity of the test to differentiate malingers [emphasis added] from genuine patients” (p. 87). Reference [1] used this as further justification for disregarding M-FAST validity studies that incorporated the SIRS. There are a number of problems with [1]’s rationale. First, [1]’s proposed criterion for the M-FAST is not entirely correct. No instrument can make the determination of malingering [4]. Instruments can only provide information regarding feigning. Therefore, the appropriate groups to compare on the M-FAST total score would consist of feigners, or potential feigners because it is a screening instrument, versus non-feigners. Since feigning is an unobservable construct, a means to measure feigning is needed, and the SIRS has demonstrated validity in identifying feigned mental illness. So, if the M-FAST cut-off score were to demonstrate agreement with the SIRS in identifying the appropriate feigning groups, it would provide validity evidence for the M-FAST cut-off. There are studies on the M-FAST which separate groups into malingering versus genuine groups. In some cases, the malingering groups would have been better labeled as feigners because motivation was not determined for the participants (e.g., Table 2 in [9]). In other cases, the methodology indicated that groups were determined to be malingering using various forms of data (e.g., [31]; see below). Those studies which identified participants as malingering can be informative in establishing the validity of the M-FAST interpretations. However, the accuracy of the malingering designation is unknown because of the variability in information that was reviewed as well as how information was weighted in making the decision. Such studies then have as many, if not more, weaknesses in identifying groups to contrast on the M-FAST total score as those studies that created groups based on the SIRS.
Reference [32] discussed the many issues in identifying “fully valid” criterions for psychological constructs in general (p. 282). From the standpoint of [32], studies where feigning groups are identified by the SIRS and then compared on the M-FAST total score might be viewed as demonstrating the construct validity of the M-FAST total score interpretations. Reference [32] indicated that “construct validation takes place when an investigator believes that his instrument reflects a particular construct, to which are attached certain meanings. The proposed interpretation generates specific testable hypotheses, which are a means of confirming or disconfirming the claim” (p. 290). This was the process undertaken in studies of the M-FAST that used the SIRS: the M-FAST total score cut-off was developed to measure the construct of potential feigning; the SIRS has demonstrated validity in identifying feigning; agreement between the SIRS feigning classification and the M-FAST cut-off then provides evidence for the M-FAST interpretation. Now, [1] described such validity findings as demonstrating “only [emphasis added] so-called convergent validity [emphasis in original]” (p. 87). The agreement in findings between the M-FAST total score and SIRS classification could be described as convergent validity, which is a theorized component of construct validity [34]. However, convergent validity is not a lesser form of validity, as [1] seemed to imply. Reference [33] describe convergent evidence as a source of evidence for test validity, with other sources being based on discriminant evidence, response processes, internal structure, and consequences associated with the test. Moreover, [33] listed test-criterion relationships under the umbrella of convergent evidence. Thus, convergent findings between the M-FAST cut-off and SIRS classification contributes to the validity base of the M-FAST interpretations, and such convergent evidence is stronger than convergent findings from studies in which the reliability and validity for determining feigning or malingering is unknown (e.g., [35]). Finally, studies where participants were categorized as malingering or not could also be considered convergent evidence, as opposed to criterion validity as [1] implied, since the reliability and validity for identifying such individuals is unlikely to be “fully valid” ([32], p. 282).

It is entirely plausible that readers would still consider studies on the M-FAST that utilized the SIRS to demonstrate criterion validity. This is understandable and not an issue from the view of the author. Whether such studies are designated as demonstrating criterion validity or convergent validity, they still contribute to the validity of the interpretations for the M-FAST total score.

C. M-FAST KGCs and PCDs

KGC and PCD methodologies are designed to have improved external validity when compared to simulation designs. Therefore, KGCs and PCDs are complementary to simulation designs. KGCs are superior over PCDs, with [4] advising against the calculation of sensitivity and specificity values from PCDs. Reference [4] indicated how “the greatest risk is false positives” in PCDs (p. 12). Now, the M-FAST was designed to be a screening instrument, which means false negatives are more problematic [36]. Consequently, sensitivity/specificity estimates for the M-FAST cut-off could still be of value when such estimates are from PCDs, but the method for categorizing groups (i.e., potential feigning/feigning vs not feigning) still needs to be thoughtfully considered.

Studies which might be best classified as PCDs [37] include [38], [31], and [35]. Reference [38] examined the performance of the M-FAST cut-off among psychiatric inpatients. A cut-off of 3 on the Personality Assessment Inventory (PAI) Malingering Index was used to establish feigning and non-feigning groups. With this criterion, the recommended M-FAST cut-off ([2], p. 12) had a sensitivity of 0.80 and a specificity of 0.85. Such values appear sufficient for a screening instrument, especially since it was used with a sample that was likely experiencing significant symptomatology. Reference [37] described the use of the Malingering Index as having “significant limitations” (p. 455). However, if one considers a cut-off of 3 on the Malingering Index to be a screener for feigning, like the M-FAST cut-off, it has demonstrated adequate differentiation at said cut-off [39]. Taking the perspective that the Malingering Index is a screener, the results from [38] are informative because the M-FAST cut-off demonstrated sufficient agreement in identifying participants who might be feigning mental illness. Reference [38] is also informative because it relied on a self-report instrument, as opposed to a structured interview (e.g., SIRS), which reduces agreement due to method variance [29]. Considering the strengths of [38] (e.g., inpatient sample), it is confusing why [1] did not utilize the results from this study.

With [38], it is possible to have some idea as to the accuracy of the identified groups. In contrast, it is not possible to know the accuracies of group identification for [31] and [35]. With respect to [31], [1] made the error of writing that criterion groups were established using only the F scale and F-K index from the MMPI. Reference [31] (p. 298) designated participants as malingering or not based on the following:

Investigators diagnosed malingering based on the observations done at the convict/detainee ward, past medical records, psychometric investigations, social survey, interview with the family and examination of the forensic records. The groups “malingering (+)” and “malingering (-)” were formed according to this diagnosis. In this way, a single blind research design was planned and the diagnosis of the clinician was accepted as the gold standard in determining the malingering ± groups.

Thus, groups were not establishing based on only MMPI validity scales, and [1]’s justification to exclude this study on such basis was in error. Since [31] relied on clinicians’ integration of multiple sources of data to identify participants who were malingering, the error rates for such classifications are unknown. Consequently, the sensitivity/specificity values from [31] should not be relied on in isolation, but in the context of other results, [31] can add to the validity base for the M-FAST cut-off. Reference [31] does have some strengths. Even though the error rate for the classification is unknown, it mirrors how malingering is detected in the real-world through using multiple sources of data. Reference [31]
also used a sample from a forensic psychiatric hospital, and 61% reported experiencing psychiatric symptoms. Therefore, it provides generalizability to settings where examinees are likely experiencing psychiatric symptoms, and given [1]’s criteria for incorporating studies in their meta-analysis, this study would have been applicable. Reference [35] had a number of similarities to [31] in that malingering was determined on the basis of clinical judgement, and the participants were experiencing psychiatric symptoms. However, unlike [31], it was less clear how it was decided which participants were malingering. This makes the application of the statistics from [35] even more questionable, but again, such statistics could be useful when considered in the context of more methodologically rigorous studies. Both [31] and [35] used translated versions of the M-FAST (i.e., Turkish and Persian). Consequently, these studies might not be generalizable to the English M-FAST, but it could be informative to compare such results to those found when using the English M-FAST. With [35], though, the translation process for the M-FAST was critiqued [37].

Finally, with respect to PCDs, [40] was published after [1]. This study used a sample of veterans undergoing inpatient monitoring for epilepsy. Of the PCDs, the M-FAST cut-off had the lowest sensitivity in [40], which is problematic for a screener. However, it still demonstrated a specificity similar to other PCDs. It is possible that a lower sensitivity was observed for the M-FAST cut-off because the Fs and FBS-r scales from the MMPI-2-RF were used in part to identify groups. Both of said scales focus on identifying feigning somatic and cognitive symptoms, and the M-FAST was not created to identify feigning of these specific symptoms. Reference [40] was considered a PCD because, even though the operating characteristics of the scales used in the study have been studied individually, the operating characteristics for these scales within a decision-model is unknown (e.g., two out of three being invalid), at least to the knowledge of the author. Nonetheless, [40] is still a valuable study given the new population within which the M-FAST was examined.

All of the KGCs to be discussed here utilized the SIRS. Given the prior discussion of the SIRS, it will be accepted as a valid means for identifying potential feigners. Two of the strongest KGCs include [20] and [9]. Such studies used the standard SIRS classification for feigning (i.e., ≥ 1 primary scale in the definite range or ≥ 3 primary scales in the probable range) and incorporated clinical samples. For example, 68% of the sample in [20] had a psychotic disorder as a primary diagnosis. Both studies found good sensitivity/specificity for the recommended M-FAST cut-off (see Table II). The sensitivity/specificity statistics for [9] were estimated using the same method from [1] because it was unclear if the statistics reported in [9] included the participants from the simulation component of the design. In addition to the SIRS criteria applied in [20] and [9], the SIRS total score can be used to classify participants as feigning if a total score >76 is obtained [25]. References [41] and [42] incorporated the SIRS total score as well as the criteria from [20] and [9] to categorize participants as feigning and not feigning. Reference [41] also examined the effectiveness of the M-FAST total score in differentiating those feigning neurocognitive symptoms, and the M-FAST did not perform sufficiently in this scenario. This is expected and does not negate the validity of the M-FAST cut-off because it was not developed to screen for feigning neurocognitive symptoms. Lastly, [43] found the weakest sensitivity for the M-FAST cut-off score. This is an interesting finding given that the SIRS was used for identifying feigning groups. Reference [43] provided some hypotheses as to why this occurred (e.g., lack of subtle items on the M-FAST).

### Table II: M-FAST KGCs and PCDs

<table>
<thead>
<tr>
<th>Reference</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Comparison</th>
<th>Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>[35]</td>
<td>0.90</td>
<td>0.87</td>
<td>C</td>
<td>PCD</td>
</tr>
<tr>
<td>[41]</td>
<td>0.83</td>
<td>0.91</td>
<td>C</td>
<td>KGC</td>
</tr>
<tr>
<td>[42]</td>
<td>0.92</td>
<td>0.72</td>
<td>C</td>
<td>KGC</td>
</tr>
<tr>
<td>[36]</td>
<td>0.86</td>
<td>0.83</td>
<td>C</td>
<td>KGC/PCD</td>
</tr>
<tr>
<td>[43]</td>
<td>0.25</td>
<td>0.95</td>
<td>NC</td>
<td>KGC</td>
</tr>
<tr>
<td>[9]</td>
<td>0.88</td>
<td>0.93</td>
<td>C</td>
<td>KGC</td>
</tr>
<tr>
<td>[31]</td>
<td>1.00</td>
<td>0.78</td>
<td>C</td>
<td>PCD</td>
</tr>
<tr>
<td>[2]</td>
<td>0.93</td>
<td>0.82*</td>
<td>C</td>
<td>KGC/PCD</td>
</tr>
<tr>
<td>[40]</td>
<td>0.53</td>
<td>0.85</td>
<td>C</td>
<td>PCD</td>
</tr>
<tr>
<td>[38]</td>
<td>0.80</td>
<td>0.85</td>
<td>C</td>
<td>PCD</td>
</tr>
<tr>
<td>[20]</td>
<td>1.00</td>
<td>0.90</td>
<td>C</td>
<td>KGC</td>
</tr>
</tbody>
</table>

Note: Sensitivity and specificity values in italics were estimated using the method from [1]. M-FAST = Miller Forensic Assessment of Symphony Test; Ref = reference; Comp = comparison group; NC = nonclinical; C = clinical; PCD = partial criterion design; KGC = known-groups comparison.

Reference [1] reported a specificity value of .83, but after recreating a 2 x 2 matrix of ns, a specificity value of .82 was obtained. Since these calculated ns were used in analyses in the companion article [55], the specificity value of .82 is reported here.

Reference [37] indicated that [2] and [36] should be considered PCDs because these studies “did not use the standard SIRS classification but a broadened criterion” (p. 455). As discussed in [36], a broadened criterion was chosen to prevent too many false negatives, which would be problematic for a screening instrument. Therefore, the feigning groups established in these studies would be consistent with the purpose of the M-FAST: a group that might be feigning and needs further evaluation vs. a group that is likely responding honestly. This rationale is similar to what was discussed regarding [38]. However, [38] relied on a single cut-score, whereas [2] and [36] required participants to have at least two elevations on SIRS scales. The requirements for the latter studies appear to be more rigorous than that of [38], and the rate of detection for two elevations of SIRS scales in the probable range has been observed [25], unlike the two elevations required in [40]. From the standpoint of [37], [2] and [36] may not meet the requirements to be considered KGCs, but these studies appear to have greater methodological strength than PCDs. Given these factors, such studies will be considered to fall in between KGCs and PCDs (i.e., labelled KGC/PCD in Table II) because coding them to either of said groups would lose important information. Both of such studies incorporated clinical samples and found adequate sensitivity and specificity values for the M-FAST total score cut-off.

The studies discussed throughout this section were designed to overcome the limitations in external validity found in simulation designs. Across the PCDs, KGCs, and KGCs/PCDs, the M-FAST cut-off was generally found to have adequate operating characteristics. Reference [1] criticized using the SIRS as a method for identifying groups to compare on the M-FAST total score due to the similarities between the two measures, but those studies which did not use the SIRS to identify feigning groups demonstrated...
sensitivity and specificity values commensurate with those studies which used the SIRS. This shows that favorable results found among KGCs which used the SIRS cannot be fully explained by shared method variance. Interestingly, the lowest sensitivity value for the M-FAST cut-off was found when the SIRS was used to identify groups [43]. All but one of the studies discussed here incorporated clinical samples, which further enhances the generalizability of the results because the M-FAST is likely to be administered in settings with examinees experiencing genuine symptoms.

Similar to the discussion for the coding of simulation studies, readers could come to different conclusions for how to best code the studies discussed throughout this section. Perhaps, it would be more reliable to simply split studies into criterion (i.e., KGC, KCD/PCD, PCD) and simulation designs, similar to what was done in [3]. This more general categorization of studies could be considered an example of low-inference coding, which is preferable [22].

III. META-ANALYSIS [3]

Reference [3] was the first meta-analysis on the M-FAST total score. To negate one of the criticisms in [1] without delay, [3] did not focus on only effect sizes: sensitivity and specificity values were pooled in meta-analytic models as well. As is common in meta-analyses of feigning measures (e.g., [24]), [3] examined effect sizes (i.e., Hedges’ g) for the M-FAST total score, which provided one method for evaluating the validity of the M-FAST total score. For a detailed discussion of results, it is recommended that the reader consult [3]. Only points to rebut [1] will be highlighted.

Reference [1] criticized the inclusion of simulation studies which did not have a clinical comparison group (e.g., [2]), but [3] included such studies in order to examine through moderator analyses whether effect sizes significantly differed between those studies which did and did not have a clinical comparison group. If studies without a clinical comparison group were not included, such analyses could not be performed. Reference [1] seemed to miss this point. Moreover, by including studies without a clinical comparison group, average specificity values could be compared for studies with and without clinical comparisons. Reference [1] ignored that such statistics were reported. Now, [1] would likely criticize the specificities found in [3] among the clinical comparison groups due to the studies which used the SIRS to identify groups. As discussed above, ignoring such studies is not sufficiently justified. Despite this discussion of specificity values from [3], sensitivity values are more important for the purpose of the M-FAST cut-off. Across all studies, the average sensitivity value for the recommended M-FAST cut-off was found to be 0.83, 95% CI [0.71, 0.91], and when sensitivity values were pooled for studies which had a clinical comparison group, an average sensitivity of 0.87, 95% CI [0.80, 0.91], was observed. These observed averages appear sufficient for a screening instrument. The average sensitivity among studies which did not have a clinical comparison group was lower than desired for a screening instrument (i.e., 0.66). This was likely due to the small number of sensitivities (k = 4) with [43] and [13] skewing the average. This lower-than-expected sensitivity could have been remedied to some extent if [3] used a more appropriate sensitivity value from [13]. That is, [13] provided an overall sensitivity (0.69) among groups feigning schizophrenia (0.90), feigning incompetency (0.25), or feigning both (0.88). Reference [3] used the overall sensitivity for these groups. In hindsight, this was inappropriate because the M-FAST total score was not designed to detect feigning incompetency, and therefore, the overall sensitivity in [13] is not a representative indicator of the M-FAST total score performance.

Moderate to large amounts of heterogeneity were observed in most analyses, even among subgroups. Consequently, the true population statistics (e.g., sensitivity) for the M-FAST studies were likely to differ from each other based on study characteristics [44]. If [3] coded studies with greater refinement, homogenous groups might have been identified, but more refined coding could have its own problems: smaller k for subgroups and/or greater subjectivity in coding. Another limitation in [3] with respect to coding was how studies were categorized to have coaching. Reference [3] used very broad criteria to label studies as having coaching. Some of these studies would have been better labeled as having a “caution” because the participants were only provided a caution to present believably or a caution regarding the presence of scales to detect feigning.

Reference [3] discussed other limitations to their results (e.g., analytic methods; [45]-[48]). Despite the limitations, [3] provided some initial estimates for the operating characteristics of the M-FAST total score.

IV. META-ANALYSIS [1]

Reference [1] is filled with hyperbole and misunderstandings regarding the M-FAST as well as feigning research designs in general. Reference [1] also ignored the limitations to their study. First, [1] excluded several studies (i.e., [7], [13], [14], [17]) because they were doctoral studies and “not easily available to scientists attempting to replicate meta-analytic calculations” (p. 87). This is a poor excuse when considering that [1] reported being affiliated with a university that likely has access to dissertation databases. The importance of retrieving data from multiple sources (e.g., journals, dissertations, books, etc.) for meta-analyses has been discussed in various meta-analytic texts [49], [50]. It would have been especially relevant for [1] to include [7] and [17] because these studies used simulators who may have been familiar with the simulation scenario, providing greater generalizability. Reference [13] provided coaching to simulators. Considering that [1] expressed concerns regarding feigners not being detected by the M-FAST cut-off, [13] would have been important to include. Reference [7] also provided simulators with the caution that “acting too crazy will be easy to spot” (p. 68). Reference [1] only had two studies to provide estimates in the subgroup “instructed malingerers, warned to avoid detection.” Therefore, including [7] could have helped to provide a more adequate estimate of the population statistic, but three studies is still quite small. What is interesting is that [1] included the doctoral dissertation [11]. So, it appears that [1] had some means to obtain dissertations. Aside from dissertations, [1] did not include [38] and [10]. Both of these studies were published in journals. From [1]’s logic then, these studies should have been easier to obtain, and these studies appeared to align with [1]’s supposed criteria for including studies.
(e.g., no SIRS identified groups). Given the strengths of these studies (see above), it is problematic that they were not included. It was already discussed above how [1]’s rationale for excluding [31] was based on an erroneous interpretation of the design. Meta-analyses need to provide the largest dataset possible in analyses to improve the adequacy of the estimates and empirically examine potential methodological influences [44], [51]. The fact that [1] did not include several relevant studies is a significant limitation and likely distorted the results.

Some of the studies that [1] included deserve discussion as well. Reference [1] described [3] as fatally flawed in part because the analogue study by [2] was included in analyses. Ironically, [1] included the same study by [2] in analyses to make comparisons among subgroups, which was the same procedure in [3]. Using [1]’s own logic, this would appear to be a fatal flaw. Based on faulty justifications, [1] reportedly excluded any studies which used the SIRS to identify groups. However, [1] included the M and SD from the group in [9] that was “suspected malingering” based on the SIRS criteria (p. 204). It is quite confusing that [1] included these statistics given the criticisms that were made of the SIRS and [3] for incorporating M-FAST studies which used the SIRS.

It was useful to estimate sensitivities and specificities from Ms and SDs when the former statistics were not provided in the primary studies. This was perhaps the greatest strength of [1], especially since [3] did not think to use this method to increase the k. However, [1] still estimated specificities when they were presented in the primary studies. This is a significant issue because three out of four of these estimates were underestimates for the specificity of the recommended M-FAST cut-off (see Table III). This is probably the most severe example of [1] misrepresenting information, especially since [1] used these underestimates to make such strong conclusions regarding the M-FAST. Reference [15] did not provide a specificity value for the M-FAST cut-off, but one author (A. Azizian) kindly provided this data upon request. This raises the question of why [1] did not attempt to make an effort to obtain such data given the conclusions that were being promulgated, but given the lack of effort to obtain dissertations for analyses or relevant articles published in journals (e.g., [38]), it is unsurprising. It is also highly problematic that [1] made firm conclusions based on only five studies, and three of the studies used translated versions of the M-FAST with two of these studies having the lowest specificities.

Reference [1] tended to present the Ns for analyses as opposed to the number of studies or values which the analyses were based (k). By presenting the Ns, [1] created the false impression that the analyses were based on large samples of data and provided adequate estimates, which was not the case. As discussed previously, [1] focused on estimated specificity values from only five studies. Moreover, in the ANOVAs that [1] performed, the data were based on small k’s. For example, the subgroup of warned malingerers was based on data from only two studies. Such small numbers of studies are unlikely to provide adequate estimates for true values. Reference [1] also essentially performed a fixed effects meta-analysis because weighting was based on only sample sizes [44]. The downfalls of fixed effects meta-analyses have been discussed in [52]. One downfall that is particularly relevant to [1] is the likelihood that the significant results were simply due to type I error. As stated in [52], “FE [fixed effects] models typically manifest a substantial Type I bias in significance tests…” (p. 275). Thus, the significant differences observed in [1] are unlikely to be true differences. Additionally, a theoretical problem with [1]’s analyses is that it assumes all the studies share the same true value, which is unlikely the case due to methodological differences.

Reference [1] wrote about the weaknesses of using simulatores, which is a legitimate criticism against the external validity of simulation designs. This is why it is important to consult findings from simulation designs as well as PCDS and KGCs because the triangulation of results would provide the strongest conclusions for the sensitivity of a feigning instrument [5]. As shown above, the M-FAST generally shows adequate operating characteristics across studies. With simulation designs, it is also important to consider the methodological characteristics of the simulation component because the design could have incorporated elements to increase the generalizability of the simulation findings (e.g., coaching).

Reference [1] emphasized the findings of [19] and [53]. In [53], M-FAST items were reviewed for their content validity. All the authors appeared to be professionals in psychiatry and/or psychology. However, there is no indication as to their expertise in feigning research. This renders many of the conclusions questionable at best, especially considering the misunderstandings in [1]. Reference [19] did provide empirical data to show how some items had poorer performance when used with trauma-exposed veterans. These findings are worthwhile to consider whenever using the M-FAST in such a population. Nonetheless, despite the poor performance of individual items, the total score was still observed to have a specificity of 0.74 among trauma-exposed veterans with probable PTSD. Considering that the M-FAST is a screening instrument, this specificity seems acceptable. Similar to the findings of [19], [2] provided hypotheses for why examinees with genuine symptoms may endorse specific items but noted that it is less likely for examinees responding genuinely to obtain an elevated total score. Thus, [19] is informative to interpretations of the M-FAST, but it does not provide damning evidence against its use.

Reference [1] criticized the lack of M-FAST studies which used severely ill psychiatric patients. This criticism is not accurate (e.g., [2], [9], [11], [15], [20], [38]). Reference [1] either excluded many studies which included psychiatric patients based on poor justifications (e.g., the study used the SIRS) or seemingly ignored such studies (e.g., [38]). Now, it is still important to continue replicating the findings of the M-FAST with psychiatric patients, but at this point, the M-FAST total score appears to operate as an adequate screening
instrument with such populations. Reference [1]’s criticism regarding the lack of studies which used severely ill psychiatric patients is also problematic because participants cannot be so severely ill that they are unable to attend to the items [2]. Consequently, as with any psychological instrument, there are limits to the degree of pathology that examinees can display to ensure valid administration.

Reference [1] discussed the potential for malingering to not be detected by the M-FAST in the context of feigning specific symptoms. The M-FAST is a broad screen for feigning psychiatric symptoms. So, it is plausible that feigning specific symptoms could lead to lower detection rates. This was even acknowledged in the M-FAST professional manual [2]: “individuals malingering specific psychiatric disorders (e.g., anxiety disorders) may not be detected by the M-FAST due to its limited coverage of certain types of symptoms” (p. 5).

Thus, [1]’s criticism of the M-FAST cut-off failing to detect specific feigned symptomatology is nothing enlightening to those who read the professional manual. Although, [13] observed that 90% of simulators who were coached in feigning schizophrenia symptoms scored above the M-FAST cut-off. Of course, there are concerns regarding the external validity of this finding. It is always possible that potential feigners may not score above the M-FAST cut-off due to the symptoms being feigned or potential coaching, which is why examiners may want to consider additional testing given other sources of information. Reference [1] seems to imply that, because the M-FAST will not have optimal performance in every scenario, it is a poor instrument. No instrument is effective for every situation. Examinees need to consider the characteristics of the examinee, the purpose of the referral, and how these factors fit with the purpose of the M-FAST and associated research evidence.

Reference [1] criticized the use of the M-FAST with populations experiencing physical and neurological injuries. Reference [1] also provided an anecdotal and concerning example of an individual being denied services based on M-FAST results alone. Such use of the M-FAST is inappropriate as well as unethical [54]. The M-FAST was designed to be a screening instrument for potentially feigning of psychiatric symptoms, and any positive results require further evaluation before concluding that an examinee is feigning or malingering. The M-FAST was not designed to detect feigning physical or neurological injuries. Furthermore, research has not supported that the M-FAST cut-off can adequately detect potential feigning of symptoms associated with physiological injuries or neurological conditions [41], [40]. However, [41] and [40] found adequate levels of specificity, but additional replications are required given the limited number of studies.

Using the M-FAST to detect potential feigning of PTSD symptoms is more of questionable area. Some research has shown adequate sensitivity in detecting feigning PTSD symptoms (e.g., [35], [18]), but this was not always the case (e.g., [6]). Similarly, even though specificity is less of concern for a screening instrument, the M-FAST total score has demonstrated some adequate specificity levels (e.g., 0.74, [19]) with participants experiencing PTSD symptoms, but the total score had poor specificity in [18]. The results in [18] may be due to cultural factors, but it still shows that examinees experiencing PTSD symptoms could have a greater rate of false-positives.

Reference [1] referred to users of the M-FAST as “M-FAST psychologists” (p. 91). It is hard to imagine a psychologist who would base their professional identity on an instrument. There are appropriate and inappropriate uses of the M-FAST, and when used appropriately, the M-FAST could be a helpful tool in clinical practice. Reference [1] also deemed it “noteworthy” that the M-FAST manual ([2]) provided an acknowledgement to Richard Rogers, who developed the SIRS, “for his support and helpful comments” as if this an some enlightening piece of empirical data regarding the M-FAST (p. iii). Given the career of Richard Rogers, his support and comments should be seen as a benefit to the development of an instrument. Instead, in the context of the article, [1] seemed to present this information snidely.

Reference [1] had the opportunity to provide a thoughtful discussion on the appropriate use of the M-FAST and improve upon the analyses of [3], but instead, [1] misrepresented data and research to make the blanket conclusion that the M-FAST should not be used altogether. This could have been driven by [1]’s experiences with psychologists who misuse the M-FAST, especially given the anecdotal example that was provided. Instead of abandoning the M-FAST due to the potential for its misuse, professionals should be educated in its appropriate application, and those who do not use the M-FAST appropriately should go through a fitting remediation.

V. CONCLUSIONS

The conclusions in [1] are woefully faulty based on the literature base of the M-FAST thus far. Across various types of study designs, the M-FAST operating characteristics are generally adequate for a screening instrument. Nonetheless, there is still need for further study of the M-FAST to better understand the operating characteristics of the cut-off across study characteristics, populations, evaluation contexts, and languages. This latter point is particularly relevant because some of the lowest specificity values were observed when the M-FAST was administered in another language. Ultimately, the M-FAST can be a useful tool for clinicians to screen for potential feigning of psychiatric symptoms. Using the M-FAST for purposes other than screening is an egregious error on the part of the user.

ACKNOWLEDGEMENT

D. Detullio would like to thank T.D. Kennedy (PhD) for his helpful comments. D. Detullio would also like to thank (in alphabetical order by last name) K. Ahmadi (PhD), A. Azizian (PhD), and D. Hill (PsyD) for providing requested data.

FUNDING

D. Detullio did not receive any funding as part of this research.
CONFLICT OF INTEREST

D. Detullio does not declare any conflict of interest. The conclusions and opinions in this article do not represent the views of Western State Hospital.

REFERENCES


[45] Dahabreh IJ, Trikalinos TA, Lau J, Schmid CH. Univariate and bivariate likelihood-based meta-analysis methods performed comparably when marginal sensitivity and specificity were the targets of inference. Journal of Clinical Epidemiology. 2017; 83: 8 –17.


Detullio, D. Another perspective of the Miller Forensic Assessment of Symptoms Test – Part II: A quantitative review. *European Journal of Medical and Health Sciences*. Preprint.