PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	How Do We Assess Resilience and Grit among Internal Medicine Residents at the Mayo Clinic? A Longitudinal Validity Study Including Correlations with Medical Knowledge, Professionalism, and Clinical Performance
AUTHORS	Alahdab, Fares; Halvorsen, Andrew; Mandrekar, Jayawant; Vaa, Brianna; Montori, Victor; West, Colin; Murad, M. Hassan; Beckman, Thomas

VERSION 1 – REVIEW

REVIEWER	Mandi Musso USA
REVIEW RETURNED	07-Jul-2020

GENERAL COMMENTS	The concept of examining reliability and validity of the grit and resilience measures in medical residents is interesting and timely. However, this study has several flaws. First, the authors attempt to relate grit and resilience to burnout in the introduction and discussion without actually examining any measure of burnout. Rather, they examine performance (achievement) measures. It makes sense to talk about grit and resilience in terms of achievement in this article, as it is currently designed. However, the authors should not overstate the findings by relating this to burnout. Further, the rationale for the study design is unclear. I could not understand how the negative association between resilience and performance validates the measure. A better rationale and more coherent manuscript may have lead me to understand better why this finding was significant. Introduction: The introduction is long and unfocused. It is unclear why the constructs of grit and resilience were chosen to be examined in the same paper. A rationale for including both constructs in the same paper should be provided. The paragraph on the CD-RISC is unnecessary, as similar information is presented in the Methods section and further confuses the introduction. Methods: Regarding the ITE, it should be clarified that this is a standard measure administered to all residents across the country rather than a program-specific measure. Authors should explain why 40 ITE scores are missing. Results: The finding that a Beta of -0.02 that is significant should be reviewed. This seems like a very small number to be statistically

	significant Discussion: The association between resilience and scores on performance measures is unclear. 1) the authors offer an unsatisfactory rationale for using performance measures to validate the resilience scale. 2) The authors the negative relationship between resilience and performance validates the resilience scale; however, their argument is conjecture. The second paragraph states that “this finding might reflect resilient residents’ abilities to thrive…” How can the authors say that they are thriving if they are performing more poorly than their peers? What metric was used to assess “thriving?” The authors state that "our research should inform future interventions to decrease resident burnout and improve resident performance and well-being by using specific dimensions of the CD-RISC..." However, resilience was associated with poorer performance. Why would programs want to focus on that? There is no further rationale for this statement. In paragraph 3, the authors again conjecture that “this suggests that the medical profession selects gritty and resilient individuals, yet still manages to burn them out.” However, the authors did not examine rates of burnout among their residents at all. They are extrapolating conclusions that their data did not assess. Table 1 is quite confusing and poorly presented. The title is long and repeats in-text material.

REVIEWER	Nicola McKinley Queens University Belfast, Centre for Public Health/Department of General Surgery, Ulster Hospital Dundonald.
REVIEW RETURNED	14-Jul-2020

GENERAL COMMENTS	Interesting topic, well written, good numbers, results clearly presented. Some points to consider- feel free to ignore or amend. Introduction begins with a summary on burnout, but this isn't the topic or main focus of this study. Although resilience, grit and burnout are linked, with some evidence suggesting resilience is a protective factor for protecting against burnout etc., this is not stated/referenced. I don't know if you can score professionalism based on in-person conference attendance and evaluation completion. Is this a recognised measure of professionalism? I note that in the methods section you say this was validated in the previous studies of residents at the Mayo Clinic- but does this mean you validated their attendance or validated this as a measure of professionalism? Data analysis- Was the data normally distributed? Was the assumption of normality assessed visually? Good response rate- were incentives offered? Were only fully completed instruments included in the analysis? Need to acknowledge that is respondents self-selected to participate the study is subject to selection and response biases.

	"We are hopeful that further study of residents' resilience and grit will help to improve their quality of life." How?

REVIEWER	Susan Moffatt-Bruce Royal College of Physician and Surgeons of Canada
REVIEW RETURNED	17-Jul-2020

GENERAL COMMENTS	The paper is very well written and is very well articulated relative to impact. The importance of the paper related to physician well being is appreciable. Resilience and grit are important. It would be helpful for the authors to comment on "the how" to gain these sometimes less than innate characteristics.

VERSION 1 – AUTHOR RESPONSE

REVIEWER #1

1. The concept of examining reliability and validity of the grit and resilience measures in medical residents is interesting and timely. However, this study has several flaws. First, the authors attempt to relate grit and resilience to burnout in the introduction and discussion without actually examining any measure of burnout. Rather, they examine performance (achievement) measures. It makes sense to talk about grit and resilience in terms of achievement in this article, as it is currently designed. However, the authors should not overstate the findings by relating this to burnout. Further, the rationale for the study design is unclear. I could not understand how the negative association between resilience and performance validates the measure. A better rationale and more coherent manuscript may have lead me to understand better why this finding was significant.

Response: We agree with the need to do the following: 1) Clarify potential relationships between grit, resilience, and burnout. 2) Not to overstate our findings with respect to these potential relationships. 3) Better explain how the negative association between resilience and performance validates the measure.

Regarding potential relationships between grit, resilience, and burnout, the basic idea is that grit and resilience lie on opposite ends of the "wellness pole" in comparison to burnout. Therefore, the intuitive assumption is that strengths with respect to grit and resilience may counterbalance burnout. To better clarify this concept, we added the following sentence to the introduction: "Furthermore, there remains the need for further research on positive aspects of physician wellness – such as resilience and grit – which may serve to counterbalance burnout."

Regarding the goal to not overstate our findings with respect to these potential relationships, we agree that this is especially important given that we did not actually examine relationships between grit, resilience, and burnout. Consequently, we added the following limitation in the discussion section: "Fourth, although we implicate potential, counterbalancing interactions between grit and wellness with burnout, this remains speculative until there is further research that specifically examines interactions between performance on these scales among internal medicine residents." Additionally, we revised the following sentence in the discussion, "Overall, our research should inform future interventions to decrease resident burnout and improve resident performance and well-being, by using specific dimensions of the CD-RISC 10 and GRIT-S as roadmaps for curricular interventions," by simplifying it and removing any reference to burnout as follows: "Overall, our research should inform future interventions to improve resident performance and well-being by using the CD-RISC 10 and GRIT-S as roadmaps for curricular interventions."

To better explain how the negative association between resilience and performance validates the measure, we would highlight that the current approach to validity evidence states that all validity is construct validity, with supporting evidence from the categories of content, internal

structure, relations to other variables, response process, and consequences, with the first three categories being the most widely reported. In terms of relations to variables evidence (e.g., grit and resilience as related to clinical performance), the validity evidence can be positive or negative, and it's understood that the strength of the association is not absolute (i.e., occurs by degree), and that the association allows for some speculative inference regarding its potential meaning.

Hence, we added a brief paragraph to the methods section on the modern conceptualization of validity evidence in addition to support from systematic reviews that the most commonly reported sources of validity evidence are from the categories on content, internal structure, and relations to other variables as follows: "The validity argument for this study was based on a modern approach to validity which states that all validity is construct validity, and that validity evidence is gathered from the categories of content, internal structure, relations to other variables, response process, and consequences. Content refers to relationships between an assessment's wording and the construct that it purportedly measures. Internal structure refers to the degree to which instrument items fit the underlying construct and is often reported in terms of dimensionality and reliability. Relations to other variables evidence is the relationship between scores and other variables relevant to the construct being measured, such that the relationships may be positive or negative depending on the constructs being measured. Notably, research has indicated that commonly reported categories of validity evidence among education research studies come from the categories of content, internal structure, and relations to other variables." This additional text also includes two new references.

2. Introduction: The introduction is long and unfocused. It is unclear why the constructs of grit and resilience were chosen to be examined in the same paper. A rationale for including both constructs in the same paper should be provided.

Response: We agree that the introduction is too long. In response to Reviewer #2, we substantially shortened the section on burnout, since this was not the focus of our study. The rationale for including both constructs of grit and resilience in the same paper was further addressed by adding the following to the introduction: "Furthermore, there remains the need for further research on positive aspects of physician wellness – such as resilience and grit – which may serve to counterbalance burnout."

3. The paragraph on the CD-RISC is unnecessary, as similar information is presented in the Methods section and further confuses the introduction.

Response: We agree. This paragraph was abbreviated by removing unnecessary details about the scale because this information is already presented in the methods section. However, we found the need to retain the essential facts in this paragraph that pertained to previous studies of resilience, including among healthcare workers and physicians.

4. Methods: Regarding the ITE, it should be clarified that this is a standard measure administered to all residents across the country rather than a program-specific measure. Authors should explain why 40 ITE scores are missing.

Response: Thank you for this suggestion. We added the following sentence to the methods: "The ITE is administered to all U.S. IM residents annually." The reason 40 ITE scores are missing is that preliminary residents who have already matched to other specialties do not sit for the ITE during their 1 year of IM training.

5. Results: The finding that a Beta of -0.02 that is significant should be reviewed. This seems like a very small number to be statistically significant.

Response: We reviewed the analysis for this beta of -0.02 and found that it is statistically significant as reported.

6. Discussion: The association between resilience and scores on performance measures is unclear. 1) the authors offer an unsatisfactory rationale for using performance

measures to validate the resilience scale. 2) The authors the negative relationship between resilience and performance validates the resilience scale; however, their argument is conjecture.

Response: As explained in our response to comment #1 above, one of the main categories of validity evidence is relations to other variables, which are accepted to be both positive and negative, and which is understood to require an inference that will always be based, to some extent, on conjecture. In addition to revisions based on comment #1, we also provided further rationale for using performance measures as validity evidence for resilience scale scores, by adding the following statement to the "clinical performance" heading under methods: "We selected clinical performance measures as association variables for this study, because we believed that standardized assessments of performance are among the most rigorous challenges for testing residents' resilience and grit."

7. The second paragraph states that "this finding might reflect resilient residents' abilities to thrive…" How can the authors say that they are thriving if they are performing more poorly than their peers? What metric was used to assess "thriving?

Response: This is a good question. Whether a resident thrives was not determined by their performance on one of these measures, but rather, on their overall standing within our rigorous training environment. The idea is that stumbling in one of these aspects of performance did not result in any of these residents doing poorly in the program, which, arguably, reflects that they are "thriving" with respect to achieving world-class training. To elaborate on this further, we added the following: "In other words, whether a resident thrives was not determined by their performance on one of these measures, but rather, on their overall standing within our rigorous training environment."

8. The authors state that "our research should inform future interventions to decrease resident burnout and improve resident performance and well-being by using specific dimensions of the CD-RISC..." However, resilience was associated with poorer performance. Why would programs want to focus on that? There is no further rationale for this statement.

Response: This actually sentence reads, "*Overall*, our research should inform future interventions to decrease resident burnout and improve resident performance and well-being, by using specific dimensions of the CD-RISC 10 *and GRIT-S* as roadmaps for curricular interventions." The idea is that this is the first validity study of its kind among residents, and that validity evidence for both the CD-RISC and GRIT-S suggests that these instruments may be useful in GME, both as assessments and sources of curricular content. However, in order to soften the message we have revised the sentence as follows: "Overall, our research should inform future interventions to improve resident performance and well-being by using the CD-RISC 10 and GRIT-S as roadmaps for curricular interventions."

9. In paragraph 3, the authors again conjecture that "this suggests that the medical profession selects gritty and resilient individuals, yet still manages to burn them out." However, the authors did not examine rates of burnout among their residents at all. They are extrapolating conclusions that their data did not assess.

Response: We have addressed the conjecture regarding implications of associations with burnout as noted in response to comment #1. With respect to the sentence in paragraph 3, we were describing the literature (references 7 and 12); we were not describing our own study findings.

10. Table 1 is quite confusing and poorly presented. The title is long and repeats in-text material.

Response: We agree that the title of Table 1 should be simplified. Therefore, we revised it to read as follows: "Baseline Characteristics of Internal Medicine Resident Physicians." Additionally, we placed the critical information that was previously included in the title, in a

new note section at the bottom of the table. The correct total means for age, possible, and completed surveys (NA) have been added in place of the redundant “253 (100%)” notations.

REVIEWER #2

1. Interesting topic, well written, good numbers, results clearly presented.

Response: Thank you.

2. Some points to consider- feel free to ignore or amend.

Response: Thank you. We have addressed your comments below.

3. Introduction begins with a summary on burnout, but this isn't the topic or main focus of this study. Although resilience, grit and burnout are linked, with some evidence suggesting resilience is a protective factor for protecting against burnout etc., this is not stated/referenced.

Response: We agree that the introduction section includes too much information on burnout, which is not the subject of our study. Therefore, we significantly shortened the first paragraph. We also added a statement that resilience, grit, and burnout may be linked as follows: “Furthermore, there remains the need for further research on positive aspects of physician wellness – such as resilience and grit – which may serve to counterbalance burnout.”

4. I don't know if you can score professionalism based on in-person conference attendance and evaluation completion. Is this a recognised measure of professionalism? I note that in the methods section you say this was validated in the previous studies of residents at the Mayo Clinic- but does this mean you validated their attendance or validated this as a measure of professionalism?

Response: As this reviewer notes, we previously reported conference attendance and evaluation completion as measures of professionalism (Reed et al. JAMA. 2008; 300(11):1326-33). In that study, we did not simply validate residents’ attendance/evaluation completion; we actually incorporated and reported these as main sources of validity evidence (i.e., relations to other variables) for validating a measure of professionalism among internal medicine residents.

5. Data analysis- Was the data normally distributed? Was the assumption of normality assessed visually?

Response: Yes, we inspected distributions of the data used in this study and they were normally distributed.

6. Good response rate- were incentives offered?

Response: Thank you: yes, the response rate was good. No, we did not offer incentives.

7. Were only fully completed instruments included in the analysis?

Response: Yes, only fully completed instruments were included in the analysis.

8. Need to acknowledge that is respondents self-selected to participate the study is subject to selection and response biases.

Response: We agree. Therefore, we added the following limitation: “Fifth, residents’ self-selection to participate in the study introduces the potential for selection and response biases, though it is noteworthy that the response/participation rate for this study was high.”

9. "We are hopeful that further study of residents' resilience and grit will help to improve their quality of life." How?

Response: This is a good question. To further justify this assertion, we revised it to read as follows: "Since high resiliency and grit have been correlated with positive attributes in other populations, we are hopeful that further study of these traits in residents' will help to improve their quality of life."

REVIEWER #3

1. The paper is very well written and is very well articulated relative to impact. The importance of the paper related to physician well-being is appreciable. Resilience and grit are important.

Response: Thank you.

2. It would be helpful for the authors to comment on "the how" to gain these sometimes less than innate characteristics.

Response: Thank you for this suggestion. We are hopeful that our revisions in response to Reviewers #1 and #2 have improved our manuscript in this regard.

VERSION 2 – REVIEW

REVIEWER	Mandi Musso Our Lady of the Lake Regional Medical Center United States
REVIEW RETURNED	10-Sep-2020

GENERAL COMMENTS	The authors have addressed all concerns. This manuscript is interesting and timely.

VERSION 2 – AUTHOR RESPONSE

REVIEWER #1
1. The authors have addressed all concerns. This manuscript is interesting and timely.
Response: Thank you