**Response to Reviewers**

October 1, 2021

Dear Reviewers

Thank you for the insightful comments on our manuscript, “A Retrospective Review of Concussion Symptom Reporting and Trajectory Data across the State of Hawaii: Influence on Return-to-Learn”. Below we respond to the Reviewers comments that we believe have improved the manuscript.

**REVIEWER 1**

4. Methods/Approach  
Reviewer 1: Line 143: Please clarify if RTL is completed during step 3 or step 2. Table 1 how full return to school during Step 2 while the narrative says step 3.

RESPONSE: The text of Table 1 has been edited to clarify the HCAMP RTL protocol. Step 2 does not reflect the completion of RTL as school accommodations or adjustments are provided. The updated table further details the steps and indicates when RTL is completed after the removal of accommodations and/or adjustments.   
  
Line 170-171 & Table 2: Do the numbers in Table 2 reflect a single instance of injury (e.g. 1 concussion) or could it also represent test completed over the course of more than 1 suspected concussion in the same individual. It would be helpful to know the number of individuals represented in this table who sustained single versus multiple concussions.

RESPONSE: The manuscript has been revised to clarify that the number of tests completed corresponds to single injuries. Further, a table has been added to the supplemental materials to display the number of individuals, organized by sex, that composed each testing group.   
  
Lines 171-172: Please clarify if all individuals included in the study had baseline testing completed. It appears this was the case but increased clarification would help.

RESPONSE: This has been clarified in the manuscript.   
  
Lines 176-182: Thanks for including the HCAMP guidelines for post-injury test administration. I think it is important to stress that these are strictly guidelines for test administration. Adding justification in the introduction and a research question addressing a comparison between the guidelines and protocol is necessary especially if this is examined in the results. However, very rarely will everyone fit the guidelines specifically. Also, over-testing may be detrimental emotionally to the student athlete. Testing 3x within the first week post-injury seems quite high.

RESPONSE: This has been clarified in the manuscript.   
  
Lines 185-186: The authors report using the PCSS scale on ImPACT as a primary measure. Currently, ImPACT provides the option for athletes to complete the PCSS prior to and post the neurocognitive testing. I am assuming the authors only analyzed pre-testing symptoms. Please provide clarification in the manuscript.

RESPONSE: This has been clarified in the manuscript.   
  
Lines192 & Table 3: On Table 3 the authors included “numbness” within the category of “anxiety-mood symptoms”. I believe this is in error. If you look at the PCSS on ImPACT, the correct wording is “numbness/tingling” suggesting symptoms associated with a neck injury. I do not think it is appropriate to put “numbness” in an anxiety related category. For this reason, data including this result should be recalculated. This change should be reflected throughout the results.

RESPONSE: Thank you for notifying us of this error. We assigned the “numbness” symptom to the headache-migraine symptom cluster and re-ran all analyses. The necessary edits have been addressed throughout the results section.   
  
5. Results/Findings  
Reviewer 1: Lines 215-221: Did any athlete report persistent symptoms beyond the 4th day of testing. Also, were symptoms post-injury compared to symptoms reported at baseline in order to determine if there truly was a deviation from baseline level of performance. Many symptoms on the PCSS can be experienced by students without a concussion (e.g., fatigue, drowsiness, difficulty concentrating). How did the authors take this information into consideration? If they did not, please reflect this in the limitations section.

RESPONSE: This has been addressed in the limitations section.   
  
6. Discussion/Conclusions  
  
The authors appropriately addressed some of the challenges associated with measuring subjective symptoms associated with concussion. However, they did fail to address common inconsistencies often found in reporting patterns of the athletes. Athletes may choose to not report symptoms of may overinflate symptom ratings. More robust methods of symptom reported are needed that will determine if the athlete is inconsistent in symptom reporting. Can the authors suggest additional methods to measure symptom rating that may enhance reliability of symptom reporting?

RESPONSE: This has been addressed in the limitations section.

**REVIEWER 2**

Weaknesses: The authors are not able to link their results to recovery timelines – either return to learn (RTL) or return to play (RTP) – using number of post-injury tests as a proxy for recovery. Although there is some merit to this approach, it limits the conclusions that they can draw related to RTL. No data from RTL appears to be available at all – from when students returned to the classroom, to what supports or accommodations were in place, to how full return to learn was monitored, how it aligned with PCSS scores, or student academic performance. The authors also do not make a case for what the current report adds to the literature in their discussion, instead describing only how current results align with published research. In effect, the authors cannot meet their aim stated in the title of how such data could influence the “future of return to learn” given the limitations of the dataset. They may be able to discuss findings in light of RTL practices, but focus narrowly on symptom severity at the first time point injury, a well-known effect in the concussion literature (that more symptoms and greater symptom severity increases risk for more complicated recovery). As McCrory et al. (2017) states: “The strongest and most consistent predictor of slower recovery from SRC is the severity of a person’s initial symptoms in the first day, or initial few days, after injury.” Many papers and programs also currently recommend matching accommodations to symptoms, including Karen McAvoy’s Symptom Wheel as part of the REAP program, and these have been summarized in several reviews, including the ones cited by the authors and a more recent one by Kemp and O’Brien. The authors do not make a compelling case that the current report will add nuance to our understanding of how symptom reports inform care for students with concussion.

RESPONSE: we have edited the manuscript to clarify that a previously published paper from the HCAMP team, Tamura et al. (2020), uses a portion of the same sample from our study. The Tamura et al. (2020) study used a smaller sample with complete RTL outcomes. Because of this, we felt confident in our decision to use the number of ImPACT tests completed as a proxy for RTL duration as the findings aligned with the Tamura et al. (2020) results.

Weaknesses: The citation for the incidence of TBI is not typical, and doesn’t match with standard CDC or more narrow sport-related concussion estimates. It appears that a review paper of treatment for mTBI rather than epidemiological data is cited. The topics of symptom clusters and symptom differences between sexes are collapsed into a single paragraph that reads as if it should be split at “Another consistency in the literature…”

RESPONSE: the citation for epidemiology has been updated, and the topics of symptom clusters and symptom differences have been separated into two different paragraphs.

Overall though, the section could make at least a mention of how all of this information is related to SLPs, particularly given that Collins and Kontos have specifically stated that SLP therapy is appropriate and beneficial to people with concussion who exhibit the cognitive-fatigue subtype (they use a different tool and clustering than that used here – regardless, the case for the role of the SLP could be made as a part of this). On the other hand, the authors also focus a bit narrowly on cognitive symptoms in that any of the subtypes may interfere with academic performance, not just cognitive symptoms. As they state earlier in the paragraph, headaches are common triggers for cognitive symptoms.

RESPONSE: the manuscript has been edited to draw relation from the RTL process to the field of speech-language pathology.   
  
In the present state of concussion management, the authors state that “Although the successful completion of the step-wise RTP protocol presumes successful return to learn,” but this may be unfamiliar to readers and may benefit from being more explicitly explained. There is also some inconsistency in that the authors describe a 6 step RTP protocol recommended by the Berlin Guidelines (2017), but the HCAMP is based on the 2009 document (presumably the Zurich guidelines but this not cited or in the reference list). Table 1 is a bit confusing in that it seems to describe the data for the early stages rather than the protocol – how are decisions made about full return to school? How does the PCSS (the primary outcome measure presented here) fit into this protocol? Also, did the protocol remain consistent even as Zurich and then Berlin guidelines resulted in changing recommendations? Did management change over the time period of the study? The serial testing described in lines 176-182 may have been typical for 2009, but is not in line with current recommendations which rely instead on symptom reports (see Alsalaheen et al., 2017). Although this comment is shading into methods, given that the number of tests administered was used to derive groups, such change in testing approach may be important to consider.

RESPONSE: This has been addressed in the manuscript.   
  
4. Methods/Approach  
Reviewer 2: Weaknesses: Demographic data is not provided – what was the mean age of participants by sex? Were there any inclusion or exclusion criteria? Was data available on previous history of concussion or other premorbid risk factors? Overall the description of the sample is thin.

RESPONSE: The mean age of participant by sex has been added to the manuscript.   
  
The range data in Table 4 suggests that many tests were administered outside of the recommended protocol. This should at least be addressed in limitations.

RESPONSE: We believe our additions to the manuscript on individual clinical decision making address this point.   
  
Table 4 describes time between tests. Is time post injury available? Date of injury is critical to interpreting this data.

RESPONSE: We have added to our study limitations how we were unable to quantify the duration of time from injury date to date of post-injury 1 because of the inability to merge our data sets.   
  
The authors do state that groupings were based on number of tests, but they do not make it clear that this grouping is a proxy for recovery, i.e. those with more tests took longer to recover. From lines 198-203, it appears that the last ImPACT test is presumed to be the time to achieve the third step in the protocol after which RTP could commence? Is the final date of testing roughly equivalent to full RTL? Again, because the PCSS or symptom report is not explained in table 1, it is unclear how all of this aligns. If this was the final ImPACT test, presumably results were considered normal or returned to baseline? These assumptions could be more clearly detailed in the methods.

RESPONSE: This has been addressed in the Methods of the manuscript.   
  
**Although there are many comparisons covered in the statistical analysis, does the approach match the aim to examine: “(a) relative severity and frequency of symptom clusters and (b) comparison of severity and frequency of symptom clusters between the sexes”? These aims do not seem to address the formation of groups based on persistence of symptoms operationalized as number of ImPACT tests. It is not convincing that the approach with the series of ANOVAs is the best choice to answer these research questions. It does not appear to be designed to compare groups (1, 2, 3, or 4 tests) by time point, sex, and symptom cluster. For example, a regression approach may have been used to consider multiple predictors on time to last test (and presumed full RTL), or examining distribution of symptom scores across subtypes at first test across the four groups. For example, does considering clusters provide greater predictive value for longer versus shorter time to last test as compared to total symptoms? Greater explanation of how the selected approach meets the purpose of the paper is necessary.  
  
5. Results/Findings  
Reviewer 2: Weaknesses:  
Providing all but the final results as supplementary materials is unusual and will make the results section very hard to interpret for readers. In fact, I would highly recommend use of a figure to present results, or including tables in an easily accessible format. The results are hard to read and contrast across the groups, and within groups between sexes and across symptom clusters. The language appears to be copied and pasted in many places, adding to problems with readability and interpretation (for example, see lines 225-9, 242-7, and 268-272). A laundry list of comparisons is provided, but it appears that only the analysis beginning on line 315 addresses the aims and is interpretable. Examination of clusters using a similar approach would be informative.**  
  
Weaknesses:  
Unfortunately, the discussion is thin and does not interpret how results add to the literature in a meaningful way. Findings are inconsistently summarized and what discussion is present only states that current results align with previous reports rather than also explaining how it extends the field. For example, the RTL section has some nice considerations, but moves outside of what the data presented here can support. Symptoms are already recommended to be used for RTL, trajectory of clusters is not described here, and initial symptom burden is already a flag for potential protracted recovery (see author’s statement in lines 358-9). Results regarding sex are not discussed at all, with the topic being dismissed that “It was not the purpose of the present study to analyze the cause of differences in symptom reporting between the sexes.” While indeed the purpose was not to understand the cause, interpretation of the data within the context of understanding how concussion symptom clusters may manifest across sexes was an aim of this report and could have added to the literature. In lines 348-9, it is suggested that duration to resolution might be useful – did results provide a mean resolution time for those symptoms? Limitations suggest that little other data is available to add to the utility of the dataset. For example, time post injury, time to RTL and RTP would have added considerably to what could have been accomplished here.

RESPONSE: The manuscript has been revised to address this comment on the Discussion section.

**IMPORTANT COMMENTS FROM THE EDITORIAL OFFICE.**

In addition to completing the content revisions requested above please ensure you also attend to the following formatting issues:

\* Please follow up with your Co-Author, Pauline Mashima, and ask her to complete their digital author verification / disclosure form for the manuscript. This form is still outstanding for this submission. We have also resent the email with the link to the form online to Dr. Mashima for her convenience.  
  
\* After the References section of the manuscript, please provide a concise (1- or 2-sentence) description for each Supplemental file supplied.

Thank you for the insightful and helpful review.

Sincerely,

A pair of glasses

Description automatically generated with low confidence

Jim Wright, M.A., CCC-SLP

Doctoral Candidate | Clinical Supervisor

University of Oregon

Communication Disorders & Sciences



McKay Moore Sohlberg PhD, CCC-SLP

Professor

University of Oregon

Communication Disorders & Sciences