PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	An Instrument for the Assessment of Diarrheal Severity based on a Longitudinal Community-Based Study
AUTHORS	Lee, Gwenyth; Peñataro_Yori, Pablo; Paredes_Olortegui, Maribel; Caulfield, Laura; Sack, David; Fischer Walker, Christa; Black, Robert; Kosek, Margaret

VERSION 1 - REVIEW

REVIEWER	Wolf Schmidt LSHTM
REVIEW RETURNED	06-Mar-2014

GENERAL COMMENTS	I think this analysis is a nice idea and presented in a well written paper that could contribute to field studies in diarrhoea epidemiology. It might be of interest to link the risk score with the pathogens. As stool samples were collected I was wondering whether this could be added to the paper. The methods seem to be appropriate on the whole, but given my own lack of statistical expertise I would recommend running this past a statistician with experience in developing risk scores. There may be ways to make this score more parsimonious. Perhaps the authors could comment on that. My main criticism would be that this has been done on one dataset only. Are there no other datasets that can be used for validation? The authors mention this in the introduction as a possibility, but given that 7 years have passed since end of the study, one wonders whether there would have been time to acquire a second dataset from a different setting. Validation is obviously the key requirement for this score to become more widely acceptable – I would recommend exhausting all possible sources to check whether anyone has a suitable dataset already that might be used. As a stand-alone paper with just one study site, this paper may be “just another paper” of little immediate use (although of course it could encourage further study in this direction). Minor: Intro: First line: perhaps avoid the vague and possibly inappropriate term “developing world”. “Low income setting” might be an alternative that is less judgemental as to who needs to develop and accounts for within country differences in SES (as you already do in line 39). Line 46: personally I always get mixed up with the term bi-weekly. Perhaps rephrase as twice per week or whatever is meant. This is just for non native speakers like me. Methods: A brief description of the definitions of symptoms might help. For example what is meant by anorexia or malaise.

REVIEWER	Gagandeep Kang Christian Medical College, Vellore, India
REVIEW RETURNED	10-Mar-2014

GENERAL COMMENTS	The idea of classifying severity of diarrhea in the community such that studies can be made as comparable as possible is useful and long overdue and the authors are commended for taking on this challenging task. Page 6 of the file, line 27, 'included' Page 6 of the file, line 35, if data where children had two episodes of diarrhea between two anthropometric measurements were excluded, instead of taking the longer episode, what would the results be? This differs from page 7, line 50 which states 'duplicate episodes in the same month were excluded' if 'duplicate implies recurrence of diarrhea. Does the adjustment for 'seasonal variation in weight gain' (page 6, line 49) capture changes in growth velocity for the child with each episode of diarrhea? Anorexia and malaise are subjective assessments by the care giver and nausea and stomach pain require either direct reporting by the child (possible in the older children in this study) or a report by the mother which may again be subjective. Page 7 of the file, line 51 seems to bear this out and it should be discussed. Figure 1 relates to this as well, what are more objective measures that can be used in a community setting? Similarly, it would be useful to see whether the quantified lack of weight gain would remain similar or be significantly different if the more subjective symptoms were removed (page 8 of file, line 18-19). Also, line 26, what would the data be if >4 stools/day was used instead of >4 liquid stools/day? Table 3. Gaps of 'no more than 1 day' differs from separated by 72 hours in Methods. Please clarify. What is the possible explanation for most severe episodes not affecting linear growth? Anthropometry was carried out once a month which would result in a 1-3 week window post-episode, what would the ideal window be to assess short term weight loss? How was input of appropriate nutrition accounted for in the estimates on weight and height?

REVIEWER	Kristen Lewis PATH, USA
REVIEW RETURNED	18-Mar-2014

GENERAL COMMENTS	The authors address an important research question. The scoring system developed is important for measuring all-cause diarrheal disease severity in the community. While the majority of diarrheal cases identified in the community are anticipated to be mild, there

	are cases that will be severe due to parents not seeking care at the clinic or via use of outpatient treatment methods, like ORT, for their children. Thus, identification and validation of a diarrheal disease scoring system that can be utilized in the community setting to identify all-cause and all-severity diarrheal disease episodes associated with poor outcomes is important. A few points for the authors to consider are included here: •The authors mention the importance of having a scoring system that is agreed upon by the majority of researchers. The scoring system included in this manuscript is new, but has not been compared to what is considered the gold standard of diarrheal disease scoring systems—the Vesikari scoring system. In order to convince other researchers of the validity of this new scoring system, it will be important to validate it against the rotavirus-specific Vesikari scoring system. While this suggestion is related to future work, it may be an important point to mention as a follow-on to this project (i.e., compare rotavirus-positive episodes scored using the Vesikari score to rotavirus-positive episodes scored using this score to determine how well the community-based scoring system works when compared to the gold standard using measures of sensitivity, specificity, PPV, and NPV). Validation of the scoring system against the Vesikari scoring system will also help to clearly define the overlap between scoring systems in capturing severity from community and clinic cases (Figures 1 and 3). In addition, future work to be considered and mentioned in the discussion may include validating this score for use with each diarrheal pathogen in order to ensure that the scoring system can provide a common measure of severity across all-cause diarrheal pathogens. •The authors note that clinical severity scoring systems exist for relatively severe dehydrating and dysenteric disease. The Vesikari score, is focused on moderate to severe disease, not just severe disease and this would be helpful to clarify in the text. •Counting the duration of each symptom in the scoring system accounts for episodes that last for a long time, but doesn't necessarily take the intensity of symptoms into account. For example, a child could have short episode that lasts for only a couple of days, but that is intense (norovirus episodes are a good example of this). This intense episode would receive fewer points than an episode that lasted for more days, but resulted in fewer symptoms each day. It would be helpful for the authors to clarify what the impact of this is on the intended use for the scoring system. •This comment relates to point 4 of the review checklist: In order to replicate the study, one would need to understand how points were assigned for each of the severity scorecard items. For example, why does vomiting for 1-2 days receive a score of 1? Did the authors consider an analysis that looked closely at the individual cutpoints for scoring of each symptom contained in the scoring system and how that would affect the overall score and shifting of scores between the mild, moderate, and severe categories? •This comment relates to point 4 of the review checklist: In order to replicate the study, one would need to understand how the scoring system categories of "mild" (score 0-1), "moderate" (score 2-7) and "severe" (score ≥8) were identified. Also, did the authors consider determining the effect of these severity category cutpoints on scoring system accuracy?

	•Suggest adding a CONSORT-like diagram to describe how the authors arrived at n=3738 episodes with (2461 included in analysis) from 442 participants, as well as a graph showing the distribution of episodes by month/year. These details would provide the reader with a better understanding of the population that the scoring system was derived from. •Consider explaining the rationale for excluding duplicated events in the same month. •Figures 3 and 4 show the total number of acute episodes (n=3738). Suggest revising so that these figures only show the n=2461 that the severity score was built on. •It was interesting that there was an association between medium severity episodes and 9 month height and HAZ changes, but this association was not present for high-severity episodes. Does this suggest that the score is predictive of short-term adverse outcomes, but not long-term adverse outcomes? Consider adding rationale regarding the lack of association between high-severity episodes and 9 month change in height and HAZ in the discussion.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name Wolf Schmidt
Institution and Country LSHTM
Please state any competing interests or state 'None declared': None declared

I think this analysis is a nice idea and presented in a well written paper that could contribute to field studies in diarrhoea epidemiology. It might be of interest to link the risk score with the pathogens. As stool samples were collected I was wondering whether this could be added to the paper.
We thank the reviewer for this comment and agree that it would be a useful exercise. In particular, we felt that it might be interesting to note how this score, compared to others, tends to differentiate between rotavirus and non-rotavirus diarrheas. Unfortunately however rotavirus was not tested for in most samples (it was only tested for in a small subset of diarrheal samples), and we felt that a comparison of severity to other common pathogens was perhaps beyond the scope of this particular report. We have added a paragraph explaining this to the discussion section.
For the reviewer's interest we will also mention as an aside that ETEC-associated diarrhea seemed to have a severity similar to that of non-pathogen specific diarrheas, whereas Campylobacter- and Shigella-associated diarrheas both had slightly higher mean scores (non-pathogen specific mean score = 1.9, ETEC-diarrhea mean score = 2.0, Campylobacter-diarrhea mean score = 2.0, shigella-diarrhea mean score = 2.4).

The methods seem to be appropriate on the whole, but given my own lack of statistical expertise I would recommend running this past a statistician with experience in developing risk scores. There may be ways to make this score more parsimonious. Perhaps the authors could comment on that.
We appreciate the reviewer's comment and have had the manuscript reviewed by an outside statistician prior to re-submission. This statistician did not suggest any major changes to the analysis, but did suggest some minor clarifications which we have attempted to address in our edits. We attempted to make the score as parsimonious as possible by including only symptoms that improved the overall performance of the score when they were added to it and (ii) when multiple symptoms were correlated with each other and appeared lightly to relate to the same underlying phenomenon (e.g. overall duration / maximum number of stools per 24 hour period / total number of liquid and semi-liquid stools), only one was included, with the criteria for inclusion being the strongest individual association with poorer weight gain. From a statistical perspective the inclusion of a larger number of correlated symptoms might be preferable, but this consideration was weighed against the need to

keep the scoring system relatively uncomplicated.

My main criticism would be that this has been done on one dataset only. Are there no other datasets that can be used for validation? The authors mention this in the introduction as a possibility, but given that 7 years have passed since end of the study, one wonders whether there would have been time to acquire a second dataset from a different setting. Validation is obviously the key requirement for this score to become more widely acceptable – I would recommend exhausting all possible sources to check whether anyone has a suitable dataset already that might be used. As a stand-alone paper with just one study site, this paper may be "just another paper" of little immediate use (although of course it could encourage further study in this direction).

We completely agree with this comment as well, however after considerable internal discussion it was decided that this step of validation should be left for another project. In particular, it was felt that validation on another dataset would be more valuable if it could be done using data deriving from a different geographic context (preferably one with rotavirus present, as mentioned above) and ideally by another group. There are several limitations to our score, which we attempt to present honestly and in the hope that future work might be undertaken to address them.

Minor:
Intro:
First line: perhaps avoid the vague and possibly inappropriate term "developing world". "Low income setting" might be an alternative that is less judgemental as to who needs to develop and accounts for within country differences in SES (as you already do in line 39).
We agree with the reviewer and have changed the wording as suggested.

Line 46: personally I always get mixed up with the term bi-weekly. Perhaps rephrase as twice per week or whatever is meant. This is just for non native speakers like me.
We have changed this.

Methods:
A brief description of the definitions of symptoms might help. For example what is meant by anorexia or malaise.

We have added the Spanish translations used for 'malaise' and 'anorexia' and also a description of how these symptoms were described to the caregivers. For the sake of brevity we have not added these descriptions for fever, vomiting and liquid stools, but are happy to do so if the reviewer feels they would also be useful.

Reviewer: 2
Reviewer Name Gagandeep Kang
Institution and Country Christian Medical College, Vellore, India
Please state any competing interests or state 'None declared': None declared

The idea of classifying severity of diarrhea in the community such that studies can be made as comparable as possible is useful and long overdue and the authors are commended for taking on this challenging task.

Page 6 of the file, line 27, 'included'

We have corrected this.

Page 6 of the file, line 35, if data where children had two episodes of diarrhea between two anthropometric measurements were excluded, instead of taking the longer episode, what would the results be? This differs from page 7, line 50 which states 'duplicate episodes in the same month were excluded' if 'duplicate implies recurrence of diarrhea.
We have attempted to clarify the wording by changing "duplicate episodes" to "shorter episodes" on page 7. We ran a number of sensitively analyses on these omissions, some of which we have copied here for the reviewer's interest (please note that these are models may be slightly different from those

in the final draft, as they come from an earlier phase of analysis). We did not see much difference in excluding shorter episodes versus excluding episodes randomly, or by allowing duplicate episodes per month; since longer episodes were less common, we chose to keep these episodes preferentially.

1 month

change in weight (g)

-1 episode per month (randomly selected)

-episodes overlapping with anthro excluded
1 month

change in weight (g)

-1 ep per month
(longer duration preferentially selected)

-episodes overlapping with anthro excluded
1 month

change in weight (g)

-multiple episodes per month allowed
(incorrect, but presented here for comparison)

-episodes overlapping with anthro excluded

N 2461
2982
Anorexia -44.7 (83.3 -6.2)
(p=0.023) -44.2
(p=0.048) -38.0
(p=0.071)
Blood in Stool (observed by Mother) -2.7 (-61.9, 56.5)
(p=0.928) -4.7
(p=0.884) -4.8
(p=0.876)
Blood in Stool (observed by lab tech) 6.5 (-74.4, 87.3)
(p=0.876) 23.0
(p=0.606) 12.2
(p=0.755)
Fever -51.4 (-91.8, -11.0)
(p=0.013) -54.1
(p=0.015) -57.7
(p=0.006)
Nausea 0.6 (-58.5, 57.2)
(p=0.982) -0.1
(p=0.998) 3.2
(p=0.920)
Malaise -47.0 (-82.2, -11.8)
(p=0.009) -45.2
(p=0.024) -34.3
(p=0.070)

Mucus in Stool (observed by Mother) 1.0 (-41.4, 44.3)
(p=0.947) -3.1
(p=0.895) -3.3
(p=0.879)
Mucus in Stool (observed by Lab tech) 0.7 (-38.5, 37.0)
(p=0.969) 4.4
(p=0.829) -2.2
(p=0.907)
Stomach Pain -13.9 (-48.8, 21.1)
(p=0.437) -6.9
(p=0.729) 4.4
(p=0.818)
Vomiting -46.6 (-97.1, 3.9)
(p=0.071) -61.1
(p=0.034) -48.0
(p=0.080)
Four or more liquid stools in a 24/hr period -40.7 (-77.2, -4.2)
(p=0.029) -34.8
(p=0.086) -40.0
(p=0.035)
Maximum number of stools /24 hr period (continuous) -9.8* (-17.7, -1.8)
(p=0.016) -10.1
(p=0.026) -11.0
(p=0.012)
Episode Duration (continuous)
-5.5 (-12.9, 1.9)
(p=0.146) -4.5
(p=0.277) -5.2
(p=0.193)

Does the adjustment for 'seasonal variation in weight gain' (page 6, line 49) capture changes in growth velocity for the child with each episode of diarrhea?
The adjustment for seasonal variation in weight gain captures variation in weight gain that occur seasonally that would be unrelated to the episode of diarrhea, as there are some seasons when children generally gain more or less and there are also seasonal trends in diarrheal incidence. We did not however evaluate whether season might modify the effects of diarrheal symptoms on weight gain.

Anorexia and malaise are subjective assessments by the care giver and nausea and stomach pain require either direct reporting by the child (possible in the older children in this study) or a report by the mother which may again be subjective. Page 7 of the file, line 51 seems to bear this out and it should be discussed. Figure 1 relates to this as well, what are more objective measures that can be used in a community setting?
We completely agree with the reviewer, and this has also been a topic of internal discussion within our group. We have added the Spanish translation for both anorexia and malaise to the methods, as well as an explanation of how they were described to the caregivers according to the study protocols. We have also attempted to expand in the discussion on the need for relatively objective measures of severity that can be collected in this setting.

Similarly, it would be useful to see whether the quantified lack of weight gain would remain similar or be significantly different if the more subjective symptoms were removed (page 8 of file, line 18-19). Also, line 26, what would the data be if >4 stools/day was used instead of >4 liquid stools/day?
Removing anorexia (arguably the most subjective element of the severity score) from the calculation of the score lead to a very slight diminishment in model fit and also slightly diminished the overall association between the severity score and weight gain; for these reasons, anorexia was retained in the score despite its relative subjectivity (the same sensitivity analysis was performed on each score component, and we have added a mention of this to the methods). Regarding liquid stools versus semi-liquid stools or overall stools, there was a clear difference in the performance of these variables, and we have added a paragraph related to this in the discussion.

Table 3. Gaps of ‘no more than 1 day’ differs from separated by 72 hours in Methods. Please clarify.
This error in Table 3 has been corrected.

What is the possible explanation for most severe episodes not affecting linear growth?

We have expanded upon this in the discussion. Our opinion is the relatively smaller number of these episodes makes it difficult to draw a firm conclusion regarding their impact on linear growth.

Anthropometry was carried out once a month which would result in a 1-3 week window post-episode, what would the ideal window be to assess short term weight loss?
This is an interesting question. Anthropometry immediately prior to, during and after the episode would be of interest, but the large number of episodes that would have to be followed in this manner make this approach very intensive, and we feel that a month is still a short enough period that the effects of an episode that happened 1-3 weeks prior to measurement, while they may be attenuated slightly by catch-up weight gain, can still be well estimated. This might be an interesting add-on to a study where a frequent schedule of visits immediately following an episode was also desirable for some other purpose.

How was input of appropriate nutrition accounted for in the estimates on weight and height?

Unfortunately, data related to nutrition was not collected at in this cohort, so it is a limitation of this study that it cannot be readily accounted for.

Reviewer: 3
Reviewer Name Kristen Lewis
Institution and Country PATH, USA
Please state any competing interests or state ‘None declared’: None declared

The authors address an important research question. The scoring system developed is important for measuring all-cause diarrheal disease severity in the community. While the majority of diarrheal cases identified in the community are anticipated to be mild, there are cases that will be severe due to parents not seeking care at the clinic or via use of outpatient treatment methods, like ORT, for their children. Thus, identification and validation of a diarrheal disease scoring system that can be utilized in the community setting to identify all-cause and all-severity diarrheal disease episodes associated with poor outcomes is important. A few points for the authors to consider are included here:

•The authors mention the importance of having a scoring system that is agreed upon by the majority of researchers. The scoring system included in this manuscript is new, but has not been compared to what is considered the gold standard of diarrheal disease scoring systems—the Vesikari scoring system. In order to convince other researchers of the validity of this new scoring system, it will be important to validate it against the rotavirus-specific Vesikari scoring system. While this suggestion is related to future work, it may be an important point to mention as a follow-on to this project (i.e., compare rotavirus-positive episodes scored using the Vesikari score to rotavirus-positive episodes scored using this score to determine how well the community-based scoring system works when compared to the gold standard using measures of sensitivity, specificity, PPV, and NPV). Validation of the scoring system against the Vesikari scoring system will also help to clearly define the overlap between scoring systems in capturing severity from community and clinic cases (Figures 1 and 3). In addition, future work to be considered and mentioned in the discussion may include validating this score for use with each diarrheal pathogen in order to ensure that the scoring system can provide a common measure of severity across all-cause diarrheal pathogens.

We thank the reviewer for this useful comment and agree completely with the suggestions made. We have attempted to more clearly define the overlap between this scoring system and the Vesikari in a supplemental table and added a more complete discussion of the differences between these scores, which also ties into the limitations of this analysis, as well as the need for further work within the context of rotavirus specifically, in the discussion.

•The authors note that clinical severity scoring systems exist for relatively severe dehydrating and

dysenteric disease. The Vesikari score, is focused on moderate to severe disease, not just severe disease and this would be helpful to clarify in the text.
We have clarified this in the abstract and in the text.

•Counting the duration of each symptom in the scoring system accounts for episodes that last for a long time, but doesn't necessarily take the intensity of symptoms into account. For example, a child could have short episode that lasts for only a couple of days, but that is intense (norovirus episodes are a good example of this). This intense episode would receive fewer points than an episode that lasted for more days, but resulted in fewer symptoms each day. It would be helpful for the authors to clarify what the impact of this is on the intended use for the scoring system.

We agree that this is an important point and one which we discussed at length internally. We based our scoring as much as possible on what appeared most strongly associated with poorer weight gain, and this does appear to some degree to suggest that prolonged symptoms (particular those which are not just G.I. symptoms but are systemic manifestations of disease) would be classified as more severe. Taking the norovirus example, we would expect a one day episode to yield a higher number of stools in the 24 hour period (+3), one day of anorexia (+1), one day of vomiting (+1) and one day of liquid stools (+1), which would sum to an overall score of 6, or perhaps 7 if fever was also noted. This would be classified by our system as a severe episode (at the 10th percentile (without fever) or 5th percentile (with fever) for all episodes) but might still be classified as less severe than an episode of, say, Shigella, which might result in several days of fever, vomiting, and watery diarrhea.

•This comment relates to point 4 of the review checklist: In order to replicate the study, one would need to understand how points were assigned for each of the severity scorecard items. For example, why does vomiting for 1-2 days receive a score of 1? Did the authors consider an analysis that looked closely at the individual cutpoints for scoring of each symptom contained in the scoring system and how that would affect the overall score and shifting of scores between the mild, moderate, and severe categories?
We thank the reviewer for the comment, which is also an interesting one. As you suggest, various cut points for each symptom were considered, and, as mentioned above, we attempted to choose categorizations that resulted in the best model fit and a fairly uniform estimate of grams lost per point, while still keeping the scoring relatively uncomplicated and also using, where possible, the same or similar categorization to previous scores. In most cases the estimated deficit per day with the symptom was close to 20 grams (the exception being anorexia, however when other categorizations of anorexia were tested, the one we included still appeared best). We also observed that, in every case, the presence of the symptom for one day was associated with a relatively larger estimate of weight lost, and then each subsequent day with the symptom was a fairly uniform, smaller amount. This is likely to be why grouping 1-2 day of a symptom into a single category (2 days is not much worse than 1) resulted in a score with a better model fit than a score in which symptoms were categorized as (1 day =1 point, 2-3= 2 points, etc.) For our purposes it appeared that, although it was possible to build a score that resulted in a wider distribution of values within the community (i.e. fewer 'mild' episodes and more 'moderate' or 'severe' ones), this was less strongly associated with weight gain than the one we ultimately presented. We have attempted to add this briefly to the discussion.

•This comment relates to point 4 of the review checklist: In order to replicate the study, one would need to understand how the scoring system categories of "mild" (score 0-1), "moderate" (score 2-7) and "severe" (score ≥8) were identified. Also, did the authors consider determining the effect of these severity category cutpoints on scoring system accuracy?
In response to this comment and some additional internal discussion, we have decide to shift the categories presented in the earlier draft to the following: mild (score 0), moderate (score 1-6), and severe (7+), with the rationale that a score of 0 is equivalent to an episode that meet the epidemiological definition of diarrhea, but has no systemic manifestations, and a score of 7 is equivalent to the 95th percentile within the study. These categorizations were not intended to be fixed but were based on the desire to collapse the relatively few high-score episodes within the sample into a single category, for the purposes of generating a linear growth model. We have attempted to emphasize that these cut-offs were chosen as characteristics of this particular distribution only so that they will not be taken as clinical suggestions.

•Suggest adding a CONSORT-like diagram to describe how the authors arrived at n=3738 episodes

with (2461 included in analysis) from 442 participants, as well as a graph showing the distribution of episodes by month/year. These details would provide the reader with a better understanding of the population that the scoring system was derived from.
We have added such a diagram, and, in order to limit the number of total tables and figures in the paper, changed the “example’ figure previously given as Figure 2 to a supplemental table.

•Consider explaining the rationale for excluding duplicated events in the same month.
We have clarified this in the methods.

•Figures 3 and 4 show the total number of acute episodes (n=3738). Suggest revising so that these figures only show the n=2461 that the severity score was built on.
We appreciate the comment and agree that it might be a bit simpler; however, if it is acceptable to the reviewer we would prefer to retain the total number of acute episodes, because this reflects the observed distribution of symptoms within the community. Using the n=2461 number slightly inflates the apparent prevalence of symptoms because, when multiple episodes occurred within the same month, the episode of longer duration was preferentially retained, and these longer episodes were also more likely to be associated with other symptoms.

•It was interesting that there was an association between medium severity episodes and 9 month height and HAZ changes, but this association was not present for high-severity episodes. Does this suggest that the score is predictive of short-term adverse outcomes, but not long-term adverse outcomes? Consider adding rationale regarding the lack of association between high-severity episodes and 9 month change in height and HAZ in the discussion.
We have attempted to clarify this in the discussion; our opinion is only that the number of the most severe episodes was small, and the estimate of the association between these episodes and height/HAZ is correspondingly less precise.

VERSION 2 – REVIEW

REVIEWER	Wolf Schmidt London School of Hygiene and Tropical Medicine UK
REVIEW RETURNED	12-May-2014

GENERAL COMMENTS	I think the paper is acceptable now. There are a few typos across the text that are better addressed during the proof stage.

REVIEWER	Gagandeep Kang Christian Medical College, Vellore, India Some of the authors and I are located at 2 of the 8 sites in a multi-site, multi-country project that examines the relationship between growth and development and enteric infection in children, but we do not collaborate directly and I do not believe that this affects the objectivity of the review process. Therefore again, None declared.
REVIEW RETURNED	16-May-2014

- The reviewer completed the checklist but made no further comments.