Thank you for this interesting study, which provides further insights into the developmental trajectories of ELBW (and EPT) children. The findings do underline the importance of a longterm follow-up and guidance. However, I do have some points raised that need to be considered before publication.

**Abstract**
- could you please add the statistics as well in the abstract?

**Methods**
- could you please clarify in the manuscript how one needs to interpret the partial eta squared (i.e. small, medium, large effect size).
- I presume that below average is <85, average 85-115 and above average >115? But this is not clearly specified. Also adding these cutoffs in the abstract would be beneficial for the reader unfamiliar with these cutoffs.

**Results**
- I think it would be a merit to show your main results in a figure to visualize more clearly the time and interaction effect.
- In the text you refer to Figure 2, but this one does not seem to be added. There is Figure 1, but without a reference in the text.
- In Figure 1, the largest part in each group remains 'stable', though the statistics show that in the below average and average group there is a significant decrease in intelligence scores. This is a bit confusing. Were there any outliers, or do I misinterpret it?

**Discussion/Conclusion**
- typo: p11 discussion, 61% in average group, needs to be 51% as written in results section?

- the assessors have two completely different profiles: experienced versus non/less experienced person, was the child assessed by the same person on the two different time points? I also assume the undergraduate student changed since the study duration is over 5 years?
- What about the impact/relevance of other developmental disabilities such as CP, but also ASS and ADHD which are also associated with less cognitive function and more common in preterm born children? As a clinician, I would also be interested if the same conclusions are valid for children with and without such developmental disabilities, since care is often different. In other words, would a preterm born child without any developmental disabilities and average IQ at 5 years, have an equal risk of losing intelligence scores by the age of 11 than peers with a developmental disabilities? Do you have the information of these co-morbidities?

- I do feel that the conclusions warrant some nuance ("Catch-up in cognitive functions seems unlikely"). Hypothetically, could there still be variation within each developmental trajectory group? And what is the impact of other, unstudied factors like co-morbidities of CP, ASS, ADHD,... I also see in Figure 1, that a small percentage does increase. Similarly, I would not define a child with an IQ of 112 at 5 years that has an IQ of 100 at 11 years, of having cognitive deficits?

- Elaborating a bit more on the clinical implications (or future directions) of your study would be a merit. Any specific recommendations/suggestions on how to support their development? Any evidence available that could hypothesis that this decline in intelligence could be limited in case of certain interventions? And how early do we need to be?

**REVIEWER**

Reviewer name: Dr. Jeremy Miles
Institution and Country: Google Inc
Competing interests: None

**REVIEW RETURNED**

09-Jan-2022

**GENERAL COMMENTS**

The title could be more descriptive; the paper is about ‘developmental trajectories’ but there is no mention of latent trajectories or neurocognitive development.

The paper says “Statistical analyses were carried out with IBM SPSS Statistics” but earlier it tells us that some analysis was done with Mplus. Perhaps be more explicit about what was done with each package?

There is no mention of how missing data is handled. Was there casewise deletion?

The statistical analysis section (in the methods) contains information that I would expect to see in the results e.g. table 1).

"FSIQ at 5 years was 66.8±14.0". I would avoid using ±, it is not clear if this is standard error, standard deviation or confidence interval.

I'm not convinced lack of a control group is a major limitation in this study. The study was about what happens to this group of children. A control group would be another group.

**VERSION 1 – AUTHOR RESPONSE**

Dear editors,
Thank you for the thorough reading of our manuscript, by both you and the reviewers. We are most grateful for the helpful and insightful reviews and comments. We have addressed each concern, as presented below, point-by-point. Due to the additions made to the manuscript, some other deletions have been made in order not to exceed the word limit.
With these changes, we kindly hope that you consider the manuscript acceptable for publication.

Sincerely,

Anu Haavisto, PhD
Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki
Phone +358-50-370 6001
E-mail: anu.haavisto@helsinki.fi

Formatting Amendments (where applicable):

1. Word Count

Please make sure that the word count is provided on the title page. Be sure that it does not exceed the requested count.

Comment: Word count has now been added on the Title page.

Editor in Chief Comments to Author:
A Figure showing changes in scores over time would help
Comment: A figure has been added to visualize the change in test scores (see Figure 1).

Associate Editor
Comments to the Author:
- The cohort is well described throughout the study period. Is the drop-out of patients in the follow-up at 5 and 11 years in line with other follow-up studies? Can the authors comment on this.

Comment: Drop out rates vary vastly between studies and are reported in different ways, affecting the percentages. Our study had a participation rate of 67% between 5 and 11 years, leading to a participation rate of 56% since study start (excluding those who died neonatally or before age 5 years).

Mangin et al (2017) reported an exceptionally high participation rate of 97% until age 12 years (reference 8). O'Brien et al (2004) reported a participation rate of 73% between 8- and 14–15-year follow-up, leading to an overall participation rate of 67% since study start (13). Similarly, Taylor et al (2004) reported an overall participation rate of 64% until age 14 years. Allin et al (2008) reported a participation rate of 85% between 8- and 15-year follow-up, but an overall participation rate of 31% since study start (10).

A few studies with longer follow up, e.g. Stålnacke et al (2015) reported a participation rate of 85% at 5-year follow-up, of which 74% participated at 18 years (overall participation rate of 63% at 18 years; reference 9). A lower rate was reported by Linsell et al (2017) with an overall participation rate of 42% until follow up at age 19 years (23).
Hence, our participation rate is somewhat lower than generally reported. However, lower overall rates have been reported. We added a notion about this in the Limitations-section in the end of the Discussion: “The attrition rate was 33%, somewhat higher than reported in many other studies."

- P10/25: Risk factors previously associated with outcome or that were theoretically of interest were included in a multiple regression analysis. The factors provided are indeed relevant. I was wondering why prenatal steroids were not included as a factor?
  
  **Comment:** This is a very valid point and antenatal steroid treatment has been added in the regression analysis, as well as in Table 2. The risk factor of emotional problem at 11 years was excluded as not to exceed the word count and to keep focus on the early medical risk factors.

- General: Gestational age (GA) was indicated as GW (gestational weeks). This might be a not frequently used abbreviation instead of GA? I suggest to consider to adapt this throughout the paper.
  
  **Comment:** GW has been changed to GA throughout the paper.

Yours sincerely,

Reviewer: 1
Dr. Jeremy Miles, Google Inc

**Comments to the Author**

- The title could be more descriptive; the paper is about ‘developmental trajectories’ but there is no mention of latent trajectories or neurocognitive development.
  
  **Comment:** The title was changed to be more descriptive: Latent class growth analysis identified different trajectories in cognitive development of extremely low birth weight children.

- The paper says “Statistical analyses were carried out with IBM SPSS Statistics” but earlier it tells us that some analysis was done with Mplus. Perhaps be more explicit about what was done with each package?
  
  **Comment:** Thank you for pointing this out. The latent class growth analysis was the only analysis done with the Mplus. This has now been stated more clearly in the Statistical analyses-section.

- There is no mention of how missing data is handled. Was there casewise deletion?
  The statistical analysis section (in the methods) contains information that I would expect to see in the results e.g. table 1.
  
  **Comment:**
Missing data was mentioned in the Assessments-section, but this has now been moved to the Statistical analyses-section together with a mention about missing data in the background factors:

“At the 5-year time point, there were missing data for the WPPSI-R subtests Comprehension (n=3), Vocabulary (n=5), Arithmetic (n=13) and Picture Completion (n=2), and the IQs were based on the subtests available.”...“Five children had missing data on perinatal morbidities and were excluded from the regression analyses.”

In sum, there was very little missing data. In case of missing subtest, index scores could be calculated based on the data available. This is an accepted procedure according to the test manual. Hence, there was no case wise deletion in the ANCOVAs. However, there was missing data in the background variables used in the regression analysis, as stated in Table 2. Concerning Table 1 and the LCGA model fits, these were moved from the Statistical analyses-section to the Results-section.

- “FSIQ at 5 years was 66.8±14.0”. I would avoid using ±, it is not clear if this is standard error, standard deviation or confidence interval.  
  **Comment**: This is a very valid point. ± has been changed to SD in the above mentioned section on page 9.

-I’m not convinced lack of a control group is a major limitation in this study. The study was about what happens to this group of children. A control group would be another group.  
  **Comment**: Thank you for this point of view. Since different tests were used at 5 and 11 years, this could potentially affect the results. Test norms differ between test versions e.g. because of how recently the norms have been collected (the debated Flynn-effect) and which exclusion criteria has been used when collecting the norms. Including a control group would have allowed for controlling for some of this variation. We have changed the wording in this statement from “a major limitation” to “a limitation”.

Reviewer: 2
Dr. Lisa Mailleux, KU Leuven
<b>Comments to the Author</b>
Thank you for this interesting study, which provides further insights into the developmental trajectories of ELBW (and EPT) children. The findings do underline the importance of a longterm follow-up and guidance. However, I do have some points raised that need to be considered before publication.

**ABSTRACT**
- could you please add the statistics as well in the abstract?  
  **Comment**: Statistics have been added in the Abstract.

**METHODS**
- could you please clarify in the manuscript how one needs to interpret the partial eta squared (i.e. small, medium, large effect size).
Comment: This has been added on the last line of the Statistical Analyses-section as following: In $\eta^2$, .01 represents a minimal effect, .06 a medium effect, and $\geq .14$ a large effect size.

- I presume that below average is $<85$, average $85-115$ and above average $>115$? But this is not clearly specified. Also adding these cutoffs in the abstract would be beneficial for the reader unfamiliar with these cutoffs.

Comment: The cutoffs have been added in the beginning of the Results-section (in the first sentence of the section Developmental trajectories on group level) as well as in the Abstract.

RESULTS
- I think it would be a merit to show your main results in a figure to visualize more clearly the time and interaction effect.

Comment: Figure 1 has been added to visualize the change in results.

- In the text you refer to Figure 2, but this one does not seem to be added. There is Figure 1, but without a reference in the text.

Comment: Apologies for the typing error. We have added a Figure 1 (see point above) and corrected the numbering.

- In Figure 1, the largest part in each group remains 'stable', though the statistics show that in the below average and average group there is a significant decrease in intelligence scores. This is a bit confusing. Were there any outliers, or do I misinterpret it?

Comment: The figure shows the number of children with change of more than 15 scores, i.e. 1 SD, between the two time points. The analysis was done as an attempt to show how many individual children may have a clinically significant reduction in scores. Many children had a decrease in test scores of less than 15 points and would have been classified as "stable". The repeated-measures ANCOVA used continuous test scores and, hence, was a more sensitive analysis. We have now added the cut off, i.e., 15 points or 1 SD, more clearly in both the Statistical Analyses and Results sections.

DISCUSSION/CONCLUSION
- typo: p11 discussion, 61% in average group, needs to be 51% as written in results section?

Comment: Thank you for finding this error. This has been corrected on page 11.

- the assessors have two completely different profiles: experienced versus non/less experienced person, was the child assessed by the same person on the two different time points? I also assume the undergraduate student changed since the study duration is over 5 years?

Comment: The assessors did change between the two time points. We have added a statement in the end of the Assessment-section.
- What about the impact/relevance of other developmental disabilities such as CP, but also ASS and ADHD which are also associated with less cognitive function and more common in preterm born children? As a clinician, I would also be interested if the same conclusions are valid for children with and without such developmental disabilities, since care is often different. In other words, would a preterm born child without any developmental disabilities and average IQ at 5 years, have an equal risk of losing intelligence scores by the age of 11 than peers with a developmental disabilities? Do you have the information of these co-morbidities?

**Comment:** This is a very interesting comment. We have the most comprehensive data on the incidence of CP and had checked for its effect. The results did not change if children with CP at 5 years were excluded from the analysis. We have now added this result to the Results-section:

“If children with CP were excluded from the analysis, the results remained. There was a significant effect of time, \( F(1,99)=11.40, p=.001, \eta^2=.103 \), with overall test scores deteriorating by -9.8 points (95% CI: -6.5 to -13.0).”

Our data on neuropsychiatric diagnoses is scarce. Hence, we have decided to only include the perinatal risk factors in risk factor analyses, because they could be collected for the whole sample.

- I do feel that the conclusions warrant some nuance ("Catch-up in cognitive functions seems unlikely"). Hypothetically, could there still be variation within each developmental trajectory group? And what is the impact of other, unstudied factors like co-morbidities of CP, ASS, ADHD,... I also see in Figure 1, that a small percentage does increase. Similarly, I would not define a child with an IQ of 112 at 5 years that has an IQ of 100 at 11 years, of having cognitive deficits?

**Comment:** The conclusion in the Abstract was changed to “seems more uncommon” as variation within each trajectory group is expected, which is also confirmed by Figure 2. The discussion tries to emphasize the fact that these are group level results.

As can be seen in Table 3, for the average and below average groups, test scores at 11 years are at least partly 1 SD below the mean at group level. A greater difference is seen in PIQ, which is a common finding in the ELBW population. Hence, the notion that there will be an increased number of children with suboptimal intelligence at school age within the ELBW population is warranted. The analysis with change in individual children was undertaken to visualize that most children will have a stable development and it seems to be a minority of children who are at risk of not keeping up with the expected developmental pace. Minor revisions have been made in the text in order to show there is variation in the results.

- Elaborating a bit more on the clinical implications (or future directions) of your study would be a merit. Any specific recommendations/suggestions on how to support their development? Any evidence available that could hypothesis that this decline in intelligence could be limited in case of certain interventions? And how early do we need to be?
Comment: One important clinical implication is the need of long-term cognitive follow up. At least in Finland, some hospitals follow up this patient population until age 5 years, which is clearly not enough. This has been stated in the conclusions.

IQ is generally seen as a relatively stable construct. As stated in the discussion, we believe that improvements in early care and minimizing neonatal risks will give the best outcomes also for cognitive development:

“Further, improvements in perinatal care to minimize brain damage and other major neonatal morbidities, optimize nutrition and efforts to give extra support particularly to families with lower parental education are important steps in supporting cognitive and emotional development for preterm children.”

Observing cognitive deficits and describing individual cognitive weaknesses and strengths will be of immense help in the planning of individual educational interventions as well as in correct psychoeducation to the families. This has now been added in the end of the Conclusions section. The age at which different deficits will surface and, hence, should be addressed depends largely on the developmental challenge the child has.

VERSION 2 – REVIEW

| REVIEWER | Reviewer name: |
|----------|----------------|
|          | Institution and Country: |
|          | Competing interests: |
| REVIEW RETURNED |

GENERAL COMMENTS

| REVIEWER | Reviewer name: |
|----------|----------------|
|          | Institution and Country: |
|          | Competing interests: |
| REVIEW RETURNED |

GENERAL COMMENTS

| REVIEWER | Reviewer name: |
|----------|----------------|
|          | Institution and Country: |
|          | Competing interests: |
| REVIEW RETURNED |

GENERAL COMMENTS

VERSION 2 – AUTHOR RESPONSE