PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Partly-randomized, controlled study in children aged 6 to 10 to investigate motor and cognitive effects of a 9-week coordination training intervention with concurrent mental tasks.
AUTHORS	Santner, Antonia; Kopp, Martin; Federolf, Peter

VERSION 1 – REVIEW

REVIEWER	Andy Tsai National Cheng Kung University, Taiwan
REVIEW RETURNED	23-Dec-2017

GENERAL COMMENTS	1. The authors' have misrepresented much of the literature using incorrect citations to justify or support their writing. This occurs numerous times throughout the manuscript. For example: The literatures regarding the a bout of acute exercise on cognition should not be used to support long-term/chronic exercise effects on cognition. The authors should refer to the previous studies exploring the effects of chronic/long-term exercise intervention on cognition in children. 2. What is "ident" exercise? 3. Why does it make sense to compare Group A and Group B1? It seems a main drawback in this study that there is a lack of a training/experimental group (Group A2: automatization of challenging coordination exercises without concurrent mental tasks). 4. The purpose of this study is vague, and the current hypotheses are not clear enough also. Hypotheses should be more specific regarding how various interventions would differentially impact on motor and cognitive performances. 5. How to ascertain that the exercise prescription (i.e., a 50-minutes sports program per week, ten weeks) will produce the beneficial effects on motor and cognitive skills. Please justify. 6. The participants will be recruited from two elementary schools. One school was selected for the coordination-automatization-mental exercises. The other school was selected as a control group. Although the two elementary schools offer similar environmental surroundings, how to control the curriculum (e.g., physical education classes) and physical activity levels between schools? 7. The participants are children in the age range of 6 to 10 years. However, the authors will adopt the Kasel-Concentration-Task (for children aged 3-8 years) and the d2-R test (for participants aged 9 to 60 years) to assess the cognitive functions. Why? 8. It is necessary to explain why Shapiro-Wilk-Test and Leven-Test are chosen to test for normal distribution and homogeneity, respectively.

	9. Narratives about multivariate statistical methods in detail (i.e. ANOVA, ANCOVA, and MANOVA) are highly suggested. Also, how effect sizes are computed is required to explain. In addition, what is the setting value of the alpha-level for the stats?

REVIEWER	Toru Ishihara Tamagawa University Brain Science Institute, Tokyo, Japan
REVIEW RETURNED	28-Dec-2017

GENERAL COMMENTS	Comments to the Authors It was a great pleasure to read this article. The authors aim to investigate the motor and cognitive effects of a 9-week exercise intervention for children aged 6 to 10 years. The topic of this line of research is very appealing and interesting given that there have not been enough studies concerning comparisons of the effects of various types of interventions on cognitive functioning in children. Despite my enthusiasm for the present submission, I found some critical issues. All comments are made under the assumption that this protocol can be modified, but I am aware that some points may not be changeable. The authors should please modify the protocol if they can, otherwise these points should be clarified and discussed in the MS. 1. The principal shortcoming of the present study was the short duration (9 weeks) and low frequency (only one lesson per week) of the intervention. The authors mention this limitation, but they should address the appropriateness of the selected duration and frequency. To the best of my knowledge, no existing study reports cognitive benefits from an intervention of such short duration and low frequency. Indeed, such studies have not been cited in the MS. Given this lack of information, it may be necessary to revise the duration and frequency of intervention in the study. 2. Another shortcoming the authors should address concerns the study design. In this study, two schools were recruited and allocated to either an intervention group or a control group. Although the authors explain that both schools are situated in similar environmental surroundings, this procedure cannot avoid an influence of some differences between the schools. If some differences are found in the change of cognitive function between study groups, the differences could be due to variances between the schools (e.g., school curriculum, teachers, facilities, etc.) and not just the intervention. To address this shortcoming, I recommend using the cluster randomized controlled design (i.e., children in both schools are randomly allocated to intervention or control group depending on classes) or a cross-over design. 3. The authors' plan evaluates a very heterogeneous group of participants in terms of sample size. This heterogeneity weakens the statistical power of their findings. Why did the authors not plan to divide the children into groups of the same sample size? I see other such large-scale studies having several limitations concerning study design and dividing sample groups as well. Thus, I understand the authors' effort to avoid these limitations and recommend mentioning these limitations and discussing them in the MS.

	4. The authors planned to conduct two cognitive tasks (i.e., The Kasel-Concentration-Task and the d2-R test). Why do the authors evaluate similar components of cognitive function using two tasks? Both these tasks may evaluate a similar component of cognitive function that requires inhibitory control. The authors should address why they are focusing specifically on inhibitory control and utilizing these tasks. I also recommend considering the evaluation of other components of cognitive function, such as working memory, cognitive flexibility, planning, and creativity. Previous studies, intended to evaluate the relationships between cognitively challenging physical activity and various types of cognitive function, indicate a selective relationship in components of cognitive function: e.g., Schmidt, M., Jäger, K., Egger, F., Roebers, C. M., & Conzelmann, A. (2015). Cognitively Engaging Chronic Physical Activity, but Not Aerobic Exercise, Affects Executive Functions in Primary School Children: A Group-Randomized Controlled Trial. Journal of Sport and Exercise Psychology, 37(6), 575–591. doi:10.1123/jsep.2015-0069 5. The authors should perform a priori power analysis based on the study design. 6. How do the authors control for confounding variables (e.g., physical activity in general, socioeconomic status, IQ, etc.)? If the authors do not plan to control for these confounders, at the very least they should mention this as a limitation. If the authors judged it unnecessary to control for some of the confounders, please consider adding explanations as to why. 7. The authors introduce many acute studies in their MS. The present study, however, is a chronic study. The mechanisms behind the effects of exercise or other interventions on cognitive function are expected to differ between acute intervention and chronic intervention. Thus, the authors should focus on citing a greater number of chronic studies. Some important previous studies that evaluated different types of physical activity intervention on cognitive function in children using randomized controlled trials are not cited in this MS. 8. The authors should also cite some recent review articles that focused on exercise type and/or mental training and emphasize more clearly the importance, necessity, and strength of the present study: e.g., Diamond, A. & Ling, D.S. (2016). Conclusions about interventions, programs, and approaches for improving executive functions that appear justified and those that, despite much hype, do not. Developmental Cognitive Neuroscience, 18, 34-48.

VERSION 1 – AUTHOR RESPONSE

Answers for Andy Tsai

Nr.	Review Remarks	Answers
1	The authors' have misrepresented much of the literature using incorrect citations to justify or support their writing. This occurs numerous times throughout the manuscript. For example: The literatures regarding the bout of	Thank you for your careful review. While we were initially surprised by this comment, we have reassessed our review of the literature and agree with you that some misinterpretations happened. We have corrected all instances that we found

	acute exercise on cognition should not be used to support long-term/chronic exercise effects on cognition. The authors should refer to the previous studies exploring the effects of chronic/long-term exercise intervention on cognition in children.	and, as requested, added more chronic interventions studies in children. Please find all new literature references in the reference list marked in color. Please also see lines 95, 97 – 10, 102, 121 – 129, 131 – 141, 144 – 150, 152 – 156, 159 – 165, 177 – 188, 196 – 197, 208.
2	What is “ident” exercise?	With “ident” exercise we mean the “same” exercise. To avoid misunderstandings, we will replace the word “ident” with the word “same”.
3	Why does it make sense to compare Group A and Group B1? It seems a main drawback in this study that there is a lack of a training/experimental group (Group A2: automatization of challenging coordination exercises without concurrent mental tasks).	The aim of this study is to compare “Kort.X” (a training concept which consists of complex coordination exercises with concurrent mental tasks) to a standard school program educated in Austrian schools (B1). To avoid variating training conditions due to different class teachers, we involved professional trainers who are briefed to educate the same exercises and intensities in all classes. We installed the second control group (B2) to validate aspects of involving concurrent mental task training.
4	The purpose of this study is vague, and the current hypotheses are not clear enough also. Hypotheses should be more specific regarding how various interventions would differentially impact on motor and cognitive performances.	Thank you, we can see your point. We have reworded the purpose statement and hypotheses. Please see lines 218 – 230.
5	How to ascertain that the exercise prescription (i.e., a 50-minutes sports program per week, ten weeks) will produce the beneficial effects on motor and cognitive skills. Please justify.	Obviously, longer training periods and frequencies are more likely to result in higher training effects as suggested for example by Diamond/Ling (2016). However, many external boundary conditions limit how the intervention in our study could be organized: 1) The Austrian school curriculum is interrupted by frequent school holidays which last from one to nine weeks. Choosing the period from school start (September) until the beginning of Christmas holidays (December) offered the longest period for testing and training interventions (including a one-week interruption of autumn holidays). A different/longer intervention period would have implicated many further confounders to be accounted for, e.g. sport activities during holidays, seasonal activity school projects (skiing and swimming courses), limited funding availability for trainers, declining interest of school organizations etc. The chosen period was well

		thought through and offered the best compromise for researchers and school authorities. 2) The School Boards insisted on implementing intervention trainings during regular physical education classes, not as extracurricular activities. They permitted to use one lesson per week for the study intervention, while the rest of the physical education curriculum has to be fulfilled by regular teachers. The authors had to plan their study with these boundary conditions. 3) Local sport clubs offer Kort.X trainings in courses of ten lessons (one hour per week). An intervention period of 10 weeks is the typical duration of the specific program that we tested in this study. It is a valid question, whether a limited intervention time of nine weeks can provide measurable effects. For several reasons, we believe that we will be able to detect an effect if it is there: 1) The sample size is large. Even if the effect size is small, we should be able to statistically prove it. 2) Experienced Kort.X trainers report recognizable impacts after four to five Kort.X training lessons: children´s movement control and fluidity as well as the performance of mental tasks apparently improve. These observations are highly subjective but were mentioned repeatedly by various trainers.
6	The participants will be recruited from two elementary schools. One school was selected for the coordination-automatization-mental exercises. The other school was selected as a control group. Although the two elementary schools offer similar environmental surroundings, how to control the curriculum (e.g., physical education classes) and physical activity levels between schools?	This is an excellent question. First, both schools offer the same amount of physical education lessons for pupils and underlie school regulations of the Tyrolean school council - classes of each school grade have to meet pre-regulated goals. Nevertheless, actual activity levels in physical education classes depend on teacher´s inputs, experience and competence. For this reason, pre-briefed, professional trainers will educate a standardized multi-exercise program which meets school regulations and researchers are able to control homogeneous intensities throughout classes.
7	The participants are children in the age range of 6 to 10 years. However, the authors will adopt the Kasel-Concentration-Task (for children aged 3-8 years) and the d2-R test (for participants aged 9 to 60 years) to assess the cognitive functions. Why?	The authors chose attention ability tests that could be undertaken by 500 pupils within a limited time span. Since most 1st–grade students cannot write and read at school start and would hardly be able to comprehend differences in letters “d” and “p” it was not advisable to use the d2-R test for all the participants. A decision was made to use the Kasel-Concentration-Task for children in 1st- and 2nd–grade classes (age 6 to 8 years) and the d2-R test for children in 3rd- and 4th-grade classes (age 8 to 10 years). After pilot testing both tests, it was apparent that the Kasel-Concentration-Task may not be challenging enough for the age group

		(although recommended by Krampen 2007), as tested children achieved the maximum number of hits without difficulties. The conclusion to this outcome was to adapt the test by adding a second test round. By reducing the assessing time (50%) it was possible to gain further improvement opportunities for higher performing children.
8 & 9	Narratives about multivariate statistical methods in detail (i.e. ANOVA, ANCOVA, and MANOVA) are highly suggested. Also, how effect sizes are computed is required to explain. In addition, what is the setting value of the alpha-level for the stats?	Thank you. We have carefully revised our decision tree for determining the appropriate statistical tests based on whether or not the data meet the statistical prerequisites (normal distribution etc.) A detailed description of contemplable methods for data analyzing was added. Please see lines 446 – 459.

Answers for Toru Ishihara

Nr.	Review Remarks	Answers
1	The principal shortcoming of the present study was the short duration (9 weeks) and low frequency (only one lesson per week) of the intervention. The authors mention this limitation, but they should address the appropriateness of the selected duration and frequency. To the best of my knowledge, no existing study reports cognitive benefits from an intervention of such short duration and low frequency. Indeed, such studies have not been cited in the MS. Given this lack of information, it may be necessary to revise the duration and frequency of intervention in the study.	Indeed, this is a central discussion point and Reviewer 1 had a similar question. We hope it is acceptable, if we repeat our answer here: Obviously, longer training periods and frequencies are more likely to result in higher training effects as suggested for example by Diamond, /Ling (2016). However, many external boundary conditions limit how the intervention in our study could be organized: 1) The Austrian school curriculum is interrupted by frequent school holidays which last from one to nine weeks. Choosing the period from school start (September) until the beginning of Christmas holidays (December) offered the longest period for testing and training interventions (including a one-week interruption of autumn holidays). A different/longer intervention period would have implicated many further confounders to be accounted for, e.g. sport activities during holidays, seasonal activity school projects (skiing and swimming courses), limited funding availability for trainers, declining interest of school organizations etc. The chosen period was well thought through and offered the best compromise for researchers and school authorities. 2) The School Boards insisted on implementing intervention trainings during regular physical education classes, not as extracurricular activities. They permitted to use one lesson per week for the study intervention, while the rest of the physical education curriculum must be fulfilled by regular teachers. The authors had to plan their study with these boundary conditions. 3) Local sport clubs offer Kort.X trainings in courses of ten lessons (one hour per week). An intervention period of 10 weeks is the typical duration of the

		specific program that we tested in this study. It is a valid question, whether a limited intervention time of nine weeks can provide measurable effects. For several reasons, we believe that we will be able to detect an effect if it is there: 1) The sample size is large. Even if the effect size is small, we should be able to statistically prove it. 2) Experienced Kort.X trainers report recognizable impacts after four to five Kort.X training lessons: children´s movement control and fluidity as well as the performance of mental tasks apparently improve. These observations are highly subjective but were mentioned repeatedly by various trainers.
2	Another shortcoming the authors should address concerns the study design. In this study, two schools were recruited and allocated to either an intervention group or a control group. Although the authors explain that both schools are situated in similar environmental surroundings, this procedure cannot avoid an influence of some differences between the schools. If some differences are found in the change of cognitive function between study groups, the differences could be due to variances between the schools (e.g., school curriculum, teachers, facilities, etc.) and not just the intervention. To address this shortcoming, I recommend using the cluster randomized controlled design (i.e., children in both schools are randomly allocated to intervention or control group depending on classes) or a cross-over design.	This is of course a question that we expected and that we had already discussed in detail in the research group and also in meetings with the School Boards and teachers. Every approach to the study design has advantages and disadvantages. As pointed out in this question, the design we chose has shortcomings. However, it also has advantages compared to the approach proposed by the reviewer. As researchers we were worried about crossover effects between groups: When children (friends, siblings) from the same school receive different interventions, they might talk, compare, or even practice together what they have experienced in their intervention groups. By not mixing physical programs within schools cross-talk and wash-out effects can be largely avoided. Another point brought forward by the headmasters and School Boards was the possibility of competitiveness between teachers when different interventions are allocated: On the one hand, group specific exercises might also be trained in classes where they should not be practiced, and on the other hand, the school atmosphere might be negatively affect as well (e.g. when teachers do not agree with the specific intervention group to which their class is assigned for). Considering a comparatively short intervention in frequency and duration, we felt that the risk of wash-out between groups is a more serious concern than the risk that the difference between schools produces a large effect. Furthermore, the cluster design cannot correct for many other differences in preexisting conditions that might appear between groups despite of a randomization, e.g. different sports participation or other

		misbalances in potential confounding factors. A cross-over study design would certainly be the strongest research design, however, it would more than double the necessary time. Neither School Boards nor teachers were open to a cross-over design in our discussions. Hence, while being aware of the potential shortcomings of the study design, we are still convinced that it is a suitable, valid and in some aspects probably better study design than the proposed clustering approach. Please see lines 78 – 81 and 461 – 489.
3	The authors' plan evaluates a very heterogeneous group of participants in terms of sample size. This heterogeneity weakens the statistical power of their findings. Why did the authors not plan to divide the children into groups of the same sample size?	The reviewer is correct that the difference in group size is a weakness in the study design and we will add this point to the study limitations (Please see lines 78 – 81 and 461 – 489.). The main reasons for the diverging group sizes were the points listed in the response to the previous question. Nevertheless, while the difference in group size is certainly not optimal, the groups are still large enough that appropriate statistical procedures can be applied and that a high sensitivity can be expected.
4	The authors planned to conduct two cognitive tasks (i.e., The Kasel-Concentration-Task and the d2-R test). Why do the authors evaluate similar components of cognitive function using two tasks? Both these tasks may evaluate a similar component of cognitive function that requires inhibitory control. The authors should address why they are focusing specifically on inhibitory control and utilizing these tasks. I also recommend considering the evaluation of other components of cognitive function, such as working memory, cognitive flexibility, planning, and creativity. Previous studies, intended to evaluate the relationships between cognitively challenging physical activity and various types of cognitive function, indicate a selective relationship in components of cognitive function: e.g., Schmidt, M., Jäger, K., Egger, F., Roebers, C. M., & Conzelmann, A. (2015). Cognitively Engaging Chronic Physical Activity, but Not Aerobic Exercise, Affects Executive	The authors chose attention ability tests that could be undertaken by 500 pupils within a limited time span. Since most 1st–grade students cannot write and read at school start and would hardly be able to comprehend differences in letters “d” and “p” it was not advisable to use the d2-R test for all the participants. A decision was made to use the Kasel-Concentration-Task for children in 1st- and 2nd–grade classes (age 6 to 8 years) and the d2-R test for children in 3rd- and 4th-grade classes (age 8 to 10 years). We are particularly interested in the effects on attention abilities in children as they may have an important impact on academic achievements[52]. Selective attention implies the suppression of distracting surroundings and may be responsible for multiplying effects in learning abilities[53]. A study of Lan et al.[54] showed that compared to executive functions (like working memory and inhibitory control) attentional control could reliably predict achievements in language (e.g. completing sentences, reading) and mathematical skills (e.g. counting, calculating) in children.

	Functions in Primary School Children: A Group-Randomized Controlled Trial. Journal of Sport and Exercise Psychology, 37(6), 575–591. doi:10.1123/jsep.2015-0069	We want to thank you for your literature recommendations, which were helpful and have now been added to the manuscript. We recognize the opportunity of evaluating executive functions during this intervention study, however, there is only limited testing time available which cannot be prolonged due to school regulations. Hence, a decision was made to collect data on attentional aspects and possible motor performance correlations only.
5	The authors should perform a priori power analysis based on the study design.	We would like to omit the power analysis from the paper since it is based on many rough assumptions that can currently not justified well enough. Nevertheless, to demonstrate that our research design promises a relatively good sensitivity, here a simple estimation based on an assumed normal distribution and t-statistics: If we assume group sizes (after dropouts) of N1 = 200 and N2 = 80 with alpha = 0.05 and beta = 0.8, two tailed, then we will be able to detect effects down to effect sizes (Cohen's d) of 0.37. Cohen classifies effect sizes between .2 and .5 as "small", suggesting that our research design will allow detecting relatively small group differences.
6	How do the authors control for confounding variables (e.g., physical activity in general, socioeconomic status, IQ, etc.)? If the authors do not plan to control for these confounders, at the very least they should mention this as a limitation. If the authors judged it unnecessary to control for some of the confounders, please consider adding explanations as to why.	During post-testing the authors will include a questionnaire to gain supplemental information about leisure time activities (e.g. attendance in sport, dance or music classes, the amount of time watching TV or playing digital games) and children´s attitude towards the trainings intervention they received. It is not planned to include inquiries about socioeconomic status or IQ, since school boards would not allow it. Please see lines 414 – 417 and 431, 432.
7 & 8	The authors introduce many acute studies in their MS. The present study, however, is a chronic study. The mechanisms behind the effects of exercise or other interventions on cognitive function are expected to differ between acute intervention and chronic intervention. Thus, the authors should focus on citing a greater number of chronic studies. Some important previous studies that evaluated different types of physical activity intervention on cognitive function in children using randomized controlled trials are not cited in this MS. The authors should also cite some recent review articles that focused on exercise type and/or mental training and emphasize more clearly the importance, necessity, and strength of the present study: e.g., Diamond, A. & Ling, D.S. (2016). Conclusions about interventions, programs, and approaches for	Thank you for this comment and reference suggestions. Several recent reviews and chronic intervention studies have now been added in our discussion of the literature. Please also see lines 95, 97 – 10, 102, 121 – 129, 131 – 141, 144 – 150, 152 – 156, 159 – 165, 177 – 188, 196 – 197, 208 - 217.

	improving executive functions that appear justified and those that, despite much hype, do not. Developmental Cognitive Neuroscience, 18, 34-48.	

VERSION 2 – REVIEW

REVIEWER	Chia-Liang Tsai National Cheng Kung University
REVIEW RETURNED	27-Mar-2018

GENERAL COMMENTS	The authors have revised the proposal appropriately according to the reviewers' suggestions. No further comments.

REVIEWER	Toru Ishihara Tamagawa University Brain Science Institute, Japan
REVIEW RETURNED	31-Mar-2018

GENERAL COMMENTS	The authors appropriately addressed the feedback provided. The manuscript can be considered for publication.