Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

10-Nov-2015

_____
English/French

DIRECTORATE FOR EDUCATION AND SKILLS
CENTRE FOR EDUCATIONAL RESEARCH AND INNOVATION (CERI) GOVERNING BOARD

INTERVENTION AND RESEARCH PROTOCOL FOR OECD PROJECT ON ASSESSING PROGRESSION IN CREATIVE AND CRITICAL THINKING SKILLS IN EDUCATION

This paper presents the research protocol of the project of school-based assessment of creative and critical thinking skills of the OECD Centre for Educational Research and Innovation (CERI).

The project aims to develop a toolkit of educational resources that teachers can use in schools to help them foster (and assess progress in) creative and critical thinking skills: assessment rubrics, pedagogical activities and exercises, and examples of student work. The objective is to make it more visible and tangible to teachers in participating countries what it means to teach, learn and make progress in creativity and critical thinking.

This revised version provides an update of country participation and incorporates some decisions about the project design that were made since its initial publication.

Stéphan Vincent-Lancrin, Senior Analyst and Project Leader, stephan.vincent-lancrin@oecd.org

JT03386111

TABLE OF CONTENTS

Tables

**INTERVENTION AND RESEARCH PROTOCOL FOR OECD PROJECT ON
ASSESSING PROGRESSION IN CREATIVE AND CRITICAL THINKING SKILLS IN
EDUCATION**

1.   Based on the initial concept note [EDU/CERI/CD(2014)19] and the rich discussions held among the participants in the OECD project on "assessing progression in creative and critical thinking skills", this document summarises the intervention and research design of the project.

State of participation

2.   As of November 2015, the project has research teams and school networks operating in the 12 following countries/systems: Brazil, China, Finland, France, Hungary, India, Italy, the Netherlands, Slovak republic, Thailand, United Kingdom (Wales), and United States. France and the United States will have several teams participating in the project.

3.   Some other countries/systems or teams are still considering their participation in the project.

Theory of action underlying the project

4.   The aim of the project is to test and prototype a formative assessment tool to help teachers assess the creative and critical thinking skills of their students, in a traditional pedagogical setting or not. As good practice in assessment is to assess what is taught, the use of the tool will be linked to suggested pedagogical activities and exercises that give students the occasion to demonstrate their level of acquisition of these skills and whether they are making progress.

5.   There is a consensus that formal education should cultivate the creativity and critical thinking skills of students, but there is little evidence that it is done in a systematic way. One reason is that these competences or habits of mind are not assessed formally in most education systems, giving little incentives to teachers to develop them. Another, related reason is that, beyond the agreement on the broad objective, it is not clear how these skills can be made visible and tangible and articulated by teachers, students and policy makers, especially as part of the curriculum.

6.   This exploratory work is based on the assumption that the development of an international framework to assess creative and critical thinking skills in educational settings will help teachers and students develop these skills, based on a more concrete understanding of what they mean, how they can be cultivated, and how progress in these skills can be made visible in school.

7.   The language of the assessment tool will be tested and refined internationally to ensure it is teacher- and student-friendly, and a bank of pedagogies, exercises and student accomplishments will be developed to provide a proof of concept that these skills can be cultivated and documented alongside the knowledge and technical skills traditionally assessed in school.

8.   Should this first phase taking place in 2015-16 be successful, a second phase of validation of the assessment tool system-wide and of scale-up to other levels and disciplines might follow if the CERI governing board and participating countries decide to continue the work in 2017-2018. In a third phase, a summative assessment tool could be developed and items could be integrated in international tests such as the standardised tests of the OECD Programme for International Student Assessment (PISA).

School network, school levels and disciplinary areas

Schools

9.   The minimum expected size for the school and teacher sample are the following:

- 10 schools minimum per country (at least 5 at primary level and 5 at secondary level);

- 20 teachers minimum (at least 10 in primary and 10 in secondary education);

- 200 students minimum involved in the pedagogical intervention at each level (primary and secondary education);

- A control group of 200 students at each level with similar characteristics (academic level and socio-economic background) and comparable teachers (in terms of quality and experience) to those involved in the research.

10.   Countries could go for a larger sample (and it is preferable that it is the case in large countries). The number of students is mainly important for the statistical analysis of the pre- and post-test results and for any other quantitative analysis we may want to undertake. This will also ensure that we have enough exemplars of student work to get sufficient variety in terms of levels of achievement in the skills assessed. Working with 200 students in each domain and at each level would give a better grasp of whether the effects are different according to the domains. These students should be the students of the teachers participating in the study. It should also be noted that it is assumed that the 200 students will yield 200 observations (questionnaires, tests, etc.) and that they are different individuals. If it is likely not the case, countries should increase the number of students accordingly.

11.   Working with enough schools and teachers is important to maximise the chances of success of the project. It is indeed possible that, for various possible reasons, some school or teacher withdraws during the project, or that the implementation does not reach the expected quality in some schools: this is why the requirements above are really a minimum.

12.   The teachers and schools selected for the project should preferably have some previous experience in pedagogic methods aiming to foster creative and critical thinking skills. This will ease the research for the local teams. However, it is understood that there will be some variability within the international network, and participants and the Secretariat will provide an initial bank of pedagogic activities and exercises that can inspire other teams in the field, including teachers with less experience in these areas.

Table 1. Age-grade structure with transition and examination years per country

	7-8 y old	8-9 y old	9-10 old	10-11 y old	11-12 y old	12-13 y old	13-14 y old	14-15 y old
Brazil	grade 2	grade 3	grade 4	**grade 5**	grade 6	grade 7	grade 8	**grade 9**
China	grade 2	grade 3	grade 4	grade 5	grade 6	grade 7	grade 8	grade 9
France	grade 2	grade 3	grade 4	grade 5	grade 6	grade 7	grade 8	**grade 9**
Hungary	grade 2	grade 3	**grade 4**	grade 5	grade 6	grade 7	grade 8	grade 9
India	grade 2	grade 3	grade 4	grade 5	grade 6	grade 7	grade 8	**grade 9**
The Netherlands	grade 2	grade 3	grade 4	grade 5	**grade 6**	grade 7	grade 8	grade 9
Slovak Republic	grade 2	grade 3	grade 4	grade 5	grade 6	grade 7	grade 8	**grade 9**
Thailand	grade 2	grade 3	grade 4	grade 5	**grade 6**	grade 7	grade 8	**grade 9**
United States	grade 2	**grade 3**	**grade 4**	grade 5	**grade 6**	**grade 7**	**grade 8**	**grade 9**
United States	grade 2	grade 3	grade 4	grade 5	grade 6	grade 7	grade 8	grade 9
United States	grade 2	grade 3	grade 4	grade 5	grade 6	grade 7	**grade 8**	**grade 9**
Wales	grade 3	grade 4	grade 5	grade 6	grade 7	grade 8	grade 9	grade 10

Note: Grades in bold are "high stake" examination years; cells shaded are transition years.

13.   Schools and teachers in the control group should ideally have a similar structure as the group participating in the study (in terms of school, teacher, class and student numbers). Ideally, they should be offered another activity or be part of another research project and be in other schools than the ones participating in the study. It will be easier to recruit a control group if the research teams propose a light activity or intervention to the schools (in addition to the testing of the pupils). Research teams running a bigger project should select this control group in their own network.

14.   The location of the schools within countries does not matter as the project does not aim to compare the level of creative and critical skills of a country versus another. Ideally, there should however be a variety in the choice of schools, and the sample could for example include at least two schools with students from lower socio-economic backgrounds. This will allow comparing the implementation of the assessment tool (and related activities) in different contexts. The size of the schools does not matter as such, but they should preferably have an average size by the country/system's standards.

Age / Grade

15.   During the kick-off meeting of the project that was held on 29-30 April 2015, criteria to narrow down the levels and disciplines on which the group will work were set up.

16.   It was decided that in primary education the group of participants will work with students aged 8-9 or so (typically grades 2 or 3) but would avoid the transition years, that is, the first or last year of primary school, as well as, possibly, « grade 4 » (typically 10 years old), which is known among creativity psychologists as an age where there is some decrease in « creative » (or at least divergent thinking) dispositions.

Table 2.   Levels of education in which local teams will undertake the research

	Primary	Secondary
Brazil	1	1
China	1	1
Finland	1	1
France	1	1
Hungary	1	1
India	1	1
Italy		
Slovak Republic	1	1
Thailand	1	1
The Netherlands	1	1
United States (Boston)	1	0
United States (California)	1	1
United States (Michigan)	0	1
United States (Montessori)	1	1
Wales	1	1
Total	11	11

Table 3. Sub-domains of interest by country and level of education

(As of 15 June 2015)

	Primary			Secondary		
	STEM	**Arts**	**Other**	**STEM**	**Arts**	**Other**
Brazil	Maths and Science	Music and Visual arts	Interdisciplinary	Maths and Science	Music and Visual arts	Interdisciplinary
China	Maths and Science			Maths and Physics		
Finland				Science		
France	Science		Design			Design for change
Hungary	Maths		Design for change			
			Creative Partnerships			
India	Maths and Science			Maths and Science	Social studies	Interdisciplinary
Slovak Republic	Maths	Music?		Maths	Music?	
Thailand		Visual arts and Music		Maths and Science		
The Netherlands	Maths	Music or Visual arts	-	Maths, physics and chemistry	Music or Visual arts	-
United States (California)	Maths and Science	Music and Visual arts		Maths and Science	Music and Visual arts	
United States (Montessori)			Montessori			Montessori
United States (Boston)		Visual arts				
United States (Michigan)				Science		
Wales	Maths		tbd	Maths		

17.   In secondary education, it was decided to work with students aged 12-13 or so and here again avoid the transition years or years when there is some « high stake » examination for students. This should thus be around grade 7 or 8 in most countries. Given the variety of contexts, in some participating countries grade 9 is the last year of schooling for a significant share of the population.

18.   The group decided that all countries should work in both levels, when possible – even though there can be some exceptions. The objective of working with several age groups is to assess the extent to which the language and conceptual dimensions of the tool work for different ages.

19.   After having collected information about high stake examination and transition years (Table 1), it was decided to focus on 8-9 year old pupils and on 13-14 year old students. In most participating countries,

this corresponds to grades 3 and 8 (but 4 and 9 in Wales). The choice was also based on the schooling structure in England, Israel and Russia, which initially committed or considered to participate in the project. Table 2 summarises countries' or teams' plans to work in primary and/or secondary education.

Disciplines

20.   The group also agreed that all teams should, inasmuch as possible, work, at both primary and secondary levels, in at least 2 out of the 3 areas of curriculum suggested below:

- One Science, Technology, Engineering and Mathematics (STEM) discipline;

- One arts education discipline;

- One other approach (innovation study, interdisciplinary approach, etc.).

21.   It was suggested that research teams with no particular constraint or preference in this area work in mathematics education and in music or visual arts education.

22.   Table 3 presents countries' expressed intentions in terms of domain focus. Most countries intend to focus on science (6 in primary and 9 in secondary education) and mathematics (8 in primary and 7 in secondary education) as far as STEM is concerned, and on either visual arts (6 in primary and 3 in secondary education) or music (6 in primary and 4 in secondary education) as far as arts education is concerned. There are a few other interdisciplinary or project-based approaches as well and a few domains that are envisaged by only one country. As local research designs firm up, the group will have to ensure that there is enough domain commonality within the international school network to allow for comparison, and to discuss the benefits and possible downsides of this diversity.

23.   Research teams already having a planned pedagogical project or idea will provide the Secretariat with a short description of their project. This information can ultimately be shared with the whole group.

24.   In primary school, some teams may ask some teachers to apply and work with the assessment tool in two domains (e.g. maths and music) as primary school teachers often teach most if not all of the domains in most systems. This would allow seeing whether working with the proposed approach is more effective when applied to several domains than just one. This may in some contexts also simplify the work of the teachers if they use integrated pedagogies. This should however not decrease the number of (different) students and teachers participating in the project. On the contrary, interventions in multiple domains for the same students should ideally supplement rather than substitute to interventions in one domain only.

Main deliverables of the project

25.   The project is an action research expecting to produce a toolkit composed of 4 main outputs. Participating research teams and their teacher and school networks are expected 1) to contribute to build a common international framework and assessment tool that teachers can use to assess their students' creativity and critical thinking skills, with domain-specific adaptations, 2) and 3) to develop a series of exemplars of pedagogical activities and exercises making the space for these skills to be expressed, and 4) to collect a portfolio of student work.

1) Assessment tool

26.   One objective of the project is to develop, through an iterative, formative prototyping process, a robust assessment tool to assess students' skills in creativity and critical thinking. The assessment tool is

meant to be used by teachers in school as part of their usual work. The tool can be seen as a rubric, in the same spirit as the one developed by Lucas et al. (2013).

27.   The final general rubric should thus:

1. have a language that is understandable by teachers and by students;

2. reflect a conceptual framework that is understandable and consensual internationally in different settings;

3. be simple enough to be used by teachers and possibly students;

4. help teachers to design pedagogical activities and exercises related to the curriculum that give students the space to demonstrate the concerned skills;

5. give a positive description of what progress means in the selected dimensions, either for the levels in which the group will work or for a development scale covering the entire schooling time;

6. give students a reference framework for their self-assessment in these areas.

28.   The toolkit will be developed throughout the project through an incremental feedback, improvement and refinement process based on its use and discussion in the various settings of the pilot – and with the help of experts in creative and critical thinking skills.

29.   The first version of the toolkit will be proposed by the OECD Secretariat, based on a mapping and analysis of existing tools within countries and on a review of the research literature on creativity and critical thinking. The first version of the rubric may be the tool presented by Lucas et al. (2013) or a somewhat different rubric.

30.   While the group will develop one domain-general tool that works across disciplines and levels, throughout the research the domestic teams will have to apply and translate this tool to different disciplinary areas and different ages. It is possible that the development of different exercises and activities will be sufficient to make the tool domain-specific and that no change of language will be necessary for different disciplines for the assessment tool itself. However, it is also possible (and perhaps even likely) that domain-specific rubrics have to be derived from the domain-general rubric. Based on the field experience, local research teams will give feedback on the local use of the tool at different points of time, for example the end of every term, and propose adjustments of its language or focus, if necessary.

2) and 3) Portfolio of pedagogical activities and exercises:

31.   In order to make tangible how creativity and critical thinking can be taught and prepare students to their assessment, the research teams will work with teachers to develop and/or document pedagogical activities that cultivate creativity and critical thinking while teaching the official curriculum. The best activities will be shared and should be reusable by other teachers in the country, be they involved in the project or not (so ultimately a version in English and in the domestic language should be available).

32.   Exercises will also be developed in such a way that the assessment tool can be used and tested 4 to 5 times during the year. The main objective of these exercises will be to test the concepts taught and to give room for students to demonstrate their level of skills in the dimensions selected in the tool. They will also ultimately provide teachers in the participating countries with a bank of exercises that will inspire them to develop new ones.

33.   In collaboration with the OECD Secretariat, some participating teams will start devising exemplars of activities and exercises that will be shared with the whole network to inspire the development of further activities and exercises. One objective of this activity is also to determine how countries should develop their own exercises and how the pedagogical pedagogies should be documented to be easily used by other teams.

4) Portfolio of student work

34.   The research teams will also collect throughout the project examples of student achievement and responses to tests and exercises that will show progression in the acquisition of these skills and exemplify different levels of acquisition.

35.   This part of the work will be particularly important to define and refine the development scale or progression levels or steps of acquisition for the covered skills, but also to give teachers examples of what a poor, good and excellent levels look like in the grades and disciplines covered by the project.

36.   The format of these portfolios of student accomplishments still needs to be defined. It will depend on the types of pedagogical activities and exercises developed by the network. Documentation of student work will be captured digitally through pictures, video and audio clips.

Contextual data collection

37.   Research teams in the various participating countries will have to collect information about the schools, teachers and classes with which they work. While the core of the project is qualitative, this additional information will allow the group to contextualise its findings (and understand the possible differences emerging from the various countries involved). A second objective is also to get a first rough idea of the effects of teaching and assessment practices emphasising creativity and critical thinking.

38.   Information about school will include size, status (public/private, charter/academy, magnet, etc.), location, socio-economic background of students attending, level of achievement of students in comparison to other schools, participation in specific educational programmes.

39.   Information about teachers will include some biographical details, seniority, past experience with innovation projects or programmes, views on the role and assessment of creativity and critical thinking in their discipline and in their school. Information about their pedagogy should also be collected, beyond the above-mentioned pedagogical activities.

40.   Information about students will include some biographic details, information about their general achievement in school at the beginning and end of the study and in the specific disciplines studied.

41.   Students participating in the research as well as a control group not receiving the pedagogical intervention will take a common standardised pre- and post-test so that the academic levels of students in different contexts can be compared. Local research teams may have to translate these standardised tests.

42.   It was also decided to administer to the intervention and control groups the standardised creativity tests developed by Todd Lubart and his colleagues (Evaluation du Potentiel Créatif des élèves - EPoC). The advantage of these tests is that they are domain-specific and easy to translate. The marking of the tests involves some additional work for the local teams, but the results will give the group an external benchmark and also allow assessing whether the use of the tool, of the adjusted pedagogies and exercises led in enhanced creativity potential as measured by these tests (compared to the control group).

43.   An innovative (and accurate) way of documenting the pedagogies used could be done through the experience sampling method (ESM) as it would give a good idea of how teachers spend their time on different types of pedagogical activities. Another proposal was to use video to give a sense of the teaching of the participating teachers, classes and schools. The use of video may require a higher level of authorisation by local research teams and is optional. Video could also be used with some classes but not others. At the very least, research teams should document these pedagogies through interviews, observation, and other documents (e.g. lesson logs). The use of technology-enhanced pedagogies may also be a means to capture additional information about student learning and the pedagogies used, wherever possible.

44.   A sub-sample of teachers (and possibly students) should be interviewed at the beginning and end of the project. The protocol for these interviews will be discussed at a later stage.

45.   Countries can collect additional information that is relevant for their local educational system. As the project may benefit from collecting similar information in other countries, research teams should inform the Secretariat of these additional data collection plans.

46.   The pre- and post- standardised tests of academic achievement and creativity will give the group some benchmarks to compare the starting and end levels of students participating in the project. It is for example possible that differences in articulating a development scale in different countries can be attributed to students' initial knowledge and mastery in the domains covered rather than to mere cultural differences. They will also help to cast light on the relationship between traditional academic achievement, progress as assessed by student work, and creativity as measured by creativity tests. The pre- and post-tests will give information about the learning curve of the participating students, and also permit more sophisticated statistical treatment to estimate the effects of the intervention. Finally, while a control group is not necessary to contextualise the collected information and materials, the lack of a control group would not allow to ascribe any measured effect to the "creativity and critical thinking" intervention as it could correspond to the development that any similar student would have had anyway.

Work with schools, teachers and students

47.   Research teams are responsible for acquiring the research and administrative authorisations to work in and with schools, teachers and students, as appropriate in their domestic context.

48.   Research teams are expected to convene an induction meeting with the involved teachers and teams to present the research and get an initial feedback on the first version of the assessment tool. Depending on the level of proficiency and ease of involved teachers with the proposed approach, the research teams may have to provide teachers with more or less support and mentoring. On the other hand, part of their role is also to collect good ideas that were developed by teachers in the network and share them with the rest of the group.

49.   Local research teams have to work on maximising the impact and benefits of the research for the participating schools and teachers. This may involve the organisation of regular meetings with the participating teams, so that they can benefit from their respective experience and learning. This may also be done through the connection with related ongoing educational activities (such as innovation prizes, etc.).

Time demand for students

50.   The core research should take between 8-10.5 hours of students' time over the school year at most.

51.   The assessment of students should take place 4-5 times a year, that is, about every two months and when it is meaningful for a pedagogic sequence. While examples of exercises still need to be designed/collected, one could envisage that these assessments would last about 90 minutes maximum, possibly split in several sessions. It is possible that the first assessments in the year would be simpler and last for a shorter period of time. Instead of longer assessments, depending on the age of the students and context, these assessments could be administered in several brief sequences so that they remain comfortable for the students. At most, this would make a cumulative 360-450 minutes during the year (6 to 7.5 hours). The assessment could also be based on the product of student work and not be based on a formal assessment event.

52.   The pre- and post-tests with the mentioned existing psychometric test of creativity (EPoC) are taken in two parts of 20 minutes each, with one week of interval, so that would add 1.5-2 hours.

53.   The pre- and post-tests students of students' academic achievement would add an additional 2 hours maximum (depending on the selected tests).

54.   Finally, students would have to answer a questionnaire at the beginning and end of the process, which could take 40 minutes at the beginning and 20 minutes at the end.

55.   This makes altogether 10.5 to 12.5 hours of their time during the school year.

56.   However, the 6-7.5 hours of assessment should be part of their normal schooling and be a substitute (hopefully) to their traditional assessments. The assessment will indeed not just be about creative and critical thinking skills but also an assessment of learning outcomes, with the difference that it will also allow them to demonstrate their level in creative and critical thinking skills in these domains (STEM, arts, etc.). So the real extra time will be more around 6 to 8.5 hours during the school year (assuming the new assessments are a bit longer than what they used to have given that they assess more skills than the traditional ones), and not the cumulative 10.5-12.5 hours it would be if the work had no connection to their school work.

57.   A sub-sample of students should be interviewed about their perception of the process and the results of the action research, and research teams may consider using the Experience Sampling Method (ESM) in some fields if possible.

Expected role and time demand for teachers

58.   Teachers are not expected to teach a different curriculum than the official one, unless they want to, but they are expected to use the proposed rubric to think of some explicit ways to teach towards the emphasised skills/dispositions, and prepare their students for the exercises that will allow them to demonstrate these skills. As these exercises will be developed/adapted by the local teachers, depending on what they have taught, and will also assess traditional learning outcomes, teachers should be involved in their marking, perhaps with the local research team initially. By comparing the results across the international network (and within the domestic networks), shared ideas/standards about different levels in the skills assessed will gradually emerge. The collection and sharing of exemplars of student work will allow the group to gradually articulate what it means to perform at different levels for the skills/dispositions identified (e.g. "imagine", "inquire"), at their age and in their discipline. There will thus be an external moderation/comparison of the marking, probably a posteriori at the beginning, and then perhaps through the positive description of the achievement levels.

59.   In terms of time demand, here are some of the tasks for the teachers and their corresponding time estimate:

- 1 day of induction (perhaps 2 depending on the context and experience of the teachers) (8 hours);

- Marking of the exercises 4-5 times a year (this may double the time they normally take if they use traditional exercises and the requested time will depend on the size of their class and on the exercises);

- 8-10 times 2-hour follow-up sessions with the local research team (prior to and after the assessments and as needed): this could be a mix of face-to-face, skype and phone calls and would be the privileged time for feedback (16-20 hours).

60.   Other activities that can be more or less time consuming depending on their current teaching practices but are just a marginal add-on to their normal professional routine of preparing their teaching and the assessment of their students:

- Adaptation of their teaching to a greater or lesser extent based on the exemplars provided or on their ideas or on the local research team's ideas (if necessary);

- Preparation/adaptation of exercises that will make room for students to show how they perform in the chosen dimensions of the rubric (this could be done with the help of the local research team).

61.   Finally, the local research teams will need the collaboration of teachers for the purpose of collecting contextual information about the study, and teachers will be asked to:

- Answer a questionnaire at the beginning and end of the study (1.5 hours altogether);

- Document their pedagogic practices: this could be done by the observation of the research team, but also through the reporting of the teachers, ideally by ESM or other similar technique over a specific "representative" period in the year (time requested depends on the technique, but at least 1 hour);

- Some (but not all) of the teachers should be interviewed.

62.   The time demand for teachers can thus be estimated at 30 hours per year, in addition to the possible extra time and effort of adapting their teaching, of designing (or co-designing) exercises aligned with the rubric, and of marking these exercises.

Other aspects of the contextual data collection

63.   The study requires additional data collection that may or may not involve the school teachers. This should preferably be done by the research team itself, but the school teachers could equally be involved.

64.   The pre- and post-tests of students on academic achievement and on creativity will need to be administered to the students and marked according to a standard protocol.

65.   The pre- and post-tests are meant to give the group benchmarks on the comparative academic levels of the students involved in the study. These tests need not be administered by the teachers – but nothing prevents the involved teachers from administering and marking them. To avoid confusion between the assessment toolkit and the standardised tests, it is in fact preferable that the research team (rather than the teachers) administer and mark the standardised tests. The achievement tests and questionnaires will be prepared in such a way that their digitisation can be automated.

66.   The pre-test and post-test will assess the academic level and the domain-specific creativity of students at the beginning and end of the study. As mentioned above, they will inform us about how much progress students have made during the year, where they started and where they ended, and thus help us better understand differences in the contexts of the fields. They will also give us another measure of the progress made by the students (that will not be comparable to the one gained thanks to the prototyped tool). The creativity tests will have to be marked by human beings. As these tests are not related to school activities or to the curriculum, they could be marked by anyone following a certain protocol attached to the tests. For these tests, teachers should in any case not mark their own students.

67.   The current research design will provide a first tentative evaluation of the effect of working with the toolkit or in the different documented contexts using a quasi-experimental research design. However, the work remains mainly a development project and will not be able to lead to generalisable findings on impact. The sample size will likely be too small to lead to conclusive findings within countries, unless the difference between treatment and control group is spectacular, but the study will generate important information for designing a subsequent validation study.

68.   Other types of supplementary work can be performed by countries.

69.   As an internationally coordinated impact evaluation (or other type of additional research) would arguably be stronger than a mere domestic one, countries/systems/teams interested in additional optional work in this vein should inform the OECD Secretariat.

Country access to the materials and publications

70.   The results of the project, and notably the final assessment rubric and related materials, will be the sole property of the OECD. However, subject to the rights of third parties, and also subject to the rules and policies of the OECD on classification and declassification of documents, if applicable, participating countries will be granted the right to translate, copy, use and distribute, for non-commercial purposes, any documents of the toolkit produced within the framework of the project.

71.   If enough budget is available, the materials of the toolkit could be made available on a web platform that would also allow teachers and participants to communicate.

72.   Local researchers or research teams will be allowed to write and publish papers related to the local research, but only after the results of the international research have been published by the OECD.

Next meetings

73.   Two more face-to-face meetings of the group of participants are planned end of January 2016 and end of June 2016.

74.   More face-to-face meetings may be necessary, either in sub-group or for the full group, but webinars and electronic communication will be another important means of communication for the network.

Expected timeline

75.   <u>March 2015</u>: Launch. Countries sent examples of assessment tools covering some aspects of creativity and critical thinking to the OECD Secretariat. These tools were analysed by the Secretariat, along the literature, compared with others existing tools, including the rubric developed by Lucas, Claxton and Spencer (2013). A review of the related research literature is performed.

76.   29-30 April 2015: First face-to-face meeting of research coordinators. Finalisation of the design of the pedagogical intervention and of the study and first discussion of the initial assessment tool.

77.   April-November 2015: Development of the assessment tool, first attempts to link it to the disciplines investigated and to pedagogic activities and exercises. Stock taking of envisaged pedagogies and instructional activities across the international network. Development of first pedagogical tasks and examples related to instruction and local curriculum exemplifying how progress in the specified dimensions could be achieved. Face-to-face meeting with the teams developing the exemplars during the summer. Documentation of pedagogies. Development of the initial set of instruments (pre- and post-tests, questionnaires, guidelines, control files, etc.).

78.   November 2015-June 2016: Field trial in Northern hemisphere countries. Collection of information in schools and work with teachers.

79.   January 2016: Second face-to-face meeting. Feedback from the first phase. Adaptation of the language of the tool, of some dimensions, of their disciplinary or age group declinations, feedback on uses and recommendations of uses, revision of the design of the study if necessary. Collection of a second set of examples of tasks. Documentation of pedagogies.

80.   February-March 2016-December 2016: Field trial in countries with a start of the school year around February-March – mainly in the Southern hemisphere and Asia. Collection of information in schools and work with teachers.

81.   April 2016: Webinar. Feedback from the second phase of the field trial. Fine-tuning of the language of the tool, of some dimensions, of their disciplinary or age group declinations, feedback on uses and recommendations of uses, revision of the design of the study if necessary.

82.   June 2016: Third face-to-face meeting. Feedback from the third phase of the field trial. Fine tuning of the language of the tool, of some dimensions, of their disciplinary or age group declinations, feedback on uses and recommendations of uses. Collection of a second set of examples of tasks. Documentation of pedagogies. Discussion of the whole exercise and of the structure of the report.

83.   June 2016-November 2016: Preliminary data analysis. First drafts of domestic report. Complementary analyses if necessary.

84.   December 2016: Fourth face-to-face meeting.

85.   December 2016-March 2017: Delivery of all the data. All drafts of domestic reports.

86.   March 2017-December 2017: Selection and publication of the international toolkit. Drafting of final international report. Finalisation of assessment tool. Final international conference launching the materials and discussing the findings.

REFERENCES

Lucas, B., G. Claxton et E. Spencer (2013), « Progression in Student Creativity in School: First Steps Towards New Forms of Formative Assessments », *OECD Education Working Papers*, No. 86, Éditions OCDE. http://dx.doi.org/10.1787/5k4dp59msdwk-en

Table 4. Summary table of project timeline

When	What	Action	Who
Mar 2015	Launch of the project	Share examples of existing tools for the assessment of creativity and critical thinking skills.	Countries
		Analysis of the existing tools.	Secretariat
		Literature review of existing tools.	Secretariat
29 and 30 Apr 2015	1st face-to-face meeting	Finalisation of the design of the study.	Countries & Secretariat
		Discussion of the initial tool developed by Lucas et al., 2013.	Countries & Secretariat
Apr - Nov 2015	Development/selection of the tool	Stock taking of envisaged pedagogies and instructional activities across the international network.	Countries & Secretariat
		Development of first tasks and examples related to instruction and local curriculum exemplifying progress in the specified dimensions.	Countries
		A face-to-face meeting with the teams developing the examples may be necessary during the summer.	Countries
		Involvement of schools, research authorisations, choice of schools and disciplines.	Countries
Nov 2015 – Jun 2016	Field trial	Start of the field trial.	Countries & Secretariat
		Collection of information in schools.	Countries
		Adaptation and refinements of the tool and other activities and discussion through webinars	Countries & Secretariat
Jan 2016	2nd face-to-face meeting	Feedback from the first phase.	Countries & Secretariat
		Adaptation of the language of the tool, of some dimensions, of their disciplinary or age group declinations, feedback on uses and recommendations of uses, revision of the design of the study if necessary.	Countries & Secretariat
		Collection of a second set of examples of tasks.	Countries & Secretariat
		Documentation of pedagogies.	Countries & Secretariat
Feb 2015 – Dec 2016	Field trial	Start of the field trial in a second wave of countries (Southern hemisphere mainly)..	Countries & Secretariat
Apr 2016	Webinar	Feedback from the second phase of the field trial.	Countries & Secretariat
		Fine-tuning of the language of the tool, of some dimensions, of their disciplinary or age group declinations, feedback on uses and recommendations of uses, revision of the design of the study if necessary.	Countries & Secretariat
Jun 2016	3rd face-to-face meeting	Feedback from the third phase of the field trial.	Countries & Secretariat
		Fine tuning of the language of the tool, of some dimensions, of their disciplinary or age group declinations, feedback on uses and recommendations of uses.	Countries & Secretariat
		Collection of a second set of examples of tasks.	Countries
		Documentation of pedagogies.	Countries & Secretariat
		Discussion of the whole exercise and of the structure of the report.	Countries & Secretariat
Jun 2016 - Nov 2016	Draft of the domestic reports	Complementary analyses if necessary. Preliminary of the first data.	Countries & Secretariat
Dec 2016 - Mar 2017	Draft of the domestic reports	All data are reveived by the Secretariat. Final national report, including for second wave of field work.	Countries
Mar 2016 - Dec 2017	Final international report	Final assessment tool.	Secretariat
		Final international conference.	Secretariat