IMPACT OF INVENTIVE DESIGN EDUCATION THROUGH THE CORRELATION BETWEEN STUDENTS' GRADES AND INDIVIDUAL TALENT

Becattini, Niccolo; Cascini, Gaetano

Politecnico di Milano

ABSTRACT

This paper aims at assessing the impact of inventive design education on students attending a class on Methods and Tools for Systematic Innovation. The study stems from the difficulty to understand how much personal inventive talent influences the final evaluation, especially in a context where students are asked to solve open problems, as conceptual design ones. To overcome the potential bias due to the individual talent, the authors propose to determine the impact of their teaching activity by means of an ex-ante/ex-post correlation analysis. Several cohorts of students along the years have been asked to solve some design problems at the beginning of the course, when no topics have been thought yet. An adapted creativity metrics enriched to map course contents measures the students' performance at the beginning ot the class (ex-ante). These results get correlated to the students' final grades (ex-post) in order to highlight areas where teaching has a stronger impact and those where talent remains predominant.

Keywords: Design education, Creativity, Conceptual design

Contact:
Becattini, Niccolo
Politecnico di Milano
Mechanical Engineering
Italy
niccolo.becattini@polimi.it

Cite this article: Becattini, N., Cascini, G. (2019) 'Impact of Inventive Design Education through the Correlation between Students' Grades and Individual Talent', in *Proceedings of the 22nd International Conference on Engineering Design (ICED19),* Delft, The Netherlands, 5-8 August 2019. DOI:10.1017/dsi.2019.57

1 CONTEXT AND OBJECTIVE

1.1 Context

Since AY 2011/12, Politecnico di Milano introduced a class within the major of design for the MS in Mechanical Engineering and let students practice design methods and tools. Nowadays, the class is compulsory for students of the second year of the Advanced Mechanical Design track and elective for other tracks of the MS in Mechanical Engineering. The class is specifically focused on design methods and tools for systematic innovation. Its contents stress on systematic approaches and the purpose of the class is to allow the less talented students to produce creative ideas and provide those who typically face fixation and psychological inertia with additional skills to be more inventive. More in detail, the syllabus includes methods and tools to stimulate and leverage creativity for divergent thinking, as well as those supporting convergent thinking. The overarching objective of the whole class is that students start using design methods and tools in order to make their whole design activity more effective and more efficient (e.g. by generating patentable ideas which are intriguing for companies), which is beneficial both for them as design engineers and for the entities they will work for (the whole syllabus and the learning objectives are available here; https://goo.gl/eoZ8qG).

The students in this class typically face problems whose solution is open, despite constrained by the design task at hand. This kind of course, in fact, introduce the students to methods and tools to deal with inventive problems (Becattini *et al*., 2012) or, more in general, problems requiring abductive reasoning. This is significantly different from the previous classes they attend, where most of the assignments concern the solution of problems of selecting and dimensioning mechanical parts under defined loading and constraining conditions. Standard problems, whose solution is also typically closed, make easier to evaluate both the process and the outcome, as there is a reference to consider. Open problems allow multiple directions of solution to be explored with equal opportunities of success. Therefore, radically different ideas or concepts, from each other, may have equal chances to provide a satisfactory solution to the problem at hand. Within this perspective, methods to foster and drive divergence and convergence are crucial. However, the evaluation of these ideas and concepts is extremely tricky, since it is typically unclear how one generates an idea (process) without an explicit declaration about that. Moreover, the idea itself can be poorly described, which makes its interpretation more complicated. Metrics for idea evaluation can come useful for a more objective assessment of students' concepts in general and provide support for the definition of final grades, but their application is typically time consuming and requires multiple coders to have a reliable assessment, which is not sustainable with large classes of students.

The goodness of solutions and the correct application of the learned methods and tools represent a good litmus test for the effectiveness of the teaching, as higher grades would correspond to a more effective teaching (given that the assessment is fair). However, gifted students with an inventive talent could bias this perception, as it is acknowledged that some generate a brilliant idea and then move backwards to justify it, as a fake step-by-step application of a method. This clearly makes the evaluation of that student's design process meaningless, as it is a matter of proficiently using the methods and tools, not mimicking their use.

The assessment of the impact of teaching on students' inventive skills, however, can also be done by measuring changes in the self-perceived efficacy in problem solving activities as well as with questionnaires to assess how much the class changed their mindset (Becattini and Cascini, 2016). These results, despite significant to describe the changes in students' attitude towards problem solving activities, still contain potential sources of bias, as the self-perception does not necessarily correspond to reality (nevertheless it is important to underline that the measurement of self-efficacy is crucial to grab changes in personal attitude for crucial elements in design, as for instance persistence…).

Ex-ante/ex-post comparisons, as for self- efficacy, describe changes in between two different instants: before and after a specific external factor takes place in the observed context. The difference is expected to be dependent on the effect of this external factor.

The same approach appears to be promising to understand better the impact of teaching on students and to reduce the biasing effect of individual talent. However, ex-ante/ex-post measures relying just on final grades, e.g. from previous classes for the ex-ante condition, might be also biased because of the individual inventive talent of students. Moreover, grades from previous classes might be also biasing as in the other classes students typically do not deal with open problems of designing.

https://doi.org/10.1017/dsi.2019.57 Published online by Cambridge University Press

1.2 Objective

In order to overcome this limitation and to reduce the effect of this bias, **the authors propose a method to run ex-ante/ex-post comparisons for students, so that they provide tangible evidences of their inventive talent before the beginning of the course (when they've been not exposed to design methods and tools).** Nevertheless, all the ideas they produce at the beginning of the class require to be evaluated according to a tailored metrics that is strongly connected to the overall objectives of the class and to evaluation of individual exam tests, so that these results can be properly compared.

The next section describes the rationale and the characteristics of the metrics, highlighting its relevance for the syllabus, for the teaching objectives their relationships with the criteria for standard exam tests correction. Section 3 details the method of investigation and fully describe the computational pipeline to generate the numerical results to define the measurement for the ex-ante condition. Section 4 presents the results of the ex-ante vs ex-post assessment of generated ideas for a cohort of students, together with a discussion of the main findings and the limitation of this approach, together with some elements worth of interest that open opportunities for additional studies. The paper closes with a section that recaps the content and the main evidences emerged along the investigation.

2 TAILORING METRICS FOR IDEA GENERATION IN INVENTIVE DESIGN PROBLEMS

As said, for the evaluation of students final grades it's sometimes hard to distinguish the impact of design methods and tools from their inventive talent. To facilitate this evaluation and make it more objective, exam tests require students, when they use methods and tools, to make every design step they do explicit, so that the design process is documented and the application of methods and tools can be verified in terms of correctness. The degree of completion of the exam (number/extent of exercises completed) and the inventiveness of solutions also contribute to define the final grade for the students. Solutions (i.e. generated ideas) are the best candidate for the ex-ante/ex-post comparison. In fact, at the beginning students have not yet learned any method to properly apply and they have no specific assignment to complete for the ex-ante condition.

On the one hand, it is easy to find metrics for idea generation activities (e.g. Shah *et al*., 2003). On the other hand, it is definitely more complicate to find one tailored to describe ideas that aims at solving inventive problems. This means that the measurement of ideas should not just take into account their creative content, but that also the presence of a more or less sensible inventive step that was not previously conceived (e.g. making the idea patentable, as for the aims of the class). Some of these relevant aspects could be, however, captured according to typical design creativity metrics. Among them, the authors see some appears to be much more effective to measure inventive design, as they are sufficiently detailed to appreciate nuances of design change/proposals with good precision, i.e. with a sufficient range of values they can assume. Details and description below.

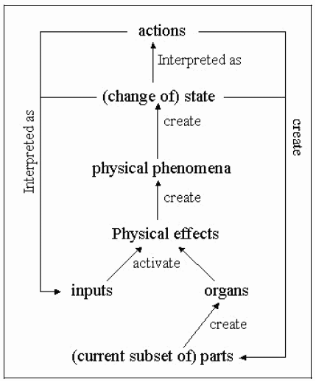


Figure 1: The sapphire model

SAPPhIRE (Figure 1- Srinivasan and Chakrabarti, 2010). It describes the cause and effect sequence which brings parts to carry out an action. Differently from FBS (Gero and Kannengiesser, 2004), which allows reformulations to be seen just at the level of Function, Behavior and Structure, this model provides a more fine-grained description of the solution's behaviour. This makes this model suitable to

https://doi.org/10.1017/dsi.2019.57 Published online by Cambridge University Press

map the degree of abstraction of the change contained in the proposed idea, with reference to a state of the art system, thus useful to recognize the degree of novelty.
Analogical Distance as characterized by Jia *et al*. (2018). This is originally for the characterization of creative stimuli as pictorial and textual examples in design experiments. This is relevant under the assumption that if it works to describe ideas expressed in text and graphics as stimuli, it should be also capable of describing the extent of exploration. Different distances depend on 1) the function of the system used as stimulus and 2) its context of application. To account for distances, functional shifts weigh more than context ones, under the assumption that it's easier to find a different context of application than allowing an existing system to deliver a new function.
Technical Plausibility as defined by Saliminamin *et al*. (2018), was originally used to measure an idea generation activity for radically innovative systems. This has to be interpreted as what accounts for "obstacles-to-market". It tries to provide an estimation of the degree of change the industry is ready to accept and/or work on to develop the solution.
Beyond the design field, there is also **Patent Law** which is a relevant source to consider, especially with reference to the objectives of the class. In fact, despite patent law does not represent a detailed metrics in itself for idea generation processes, its requirements for patentability set a clear standard of what an idea should consist of, in order to grant a patent to its inventor: the solution should be novel and non-obvious to an expert skilled in the art, moreover it should present an inventive step, meaning a concrete technological advantage.
Consistently with the above mentioned metrics, the authors propose their adaptation in order to make the following ones suitable to capture characteristics that the class aim at stimulating and training along the semester.
Purpose (→ analogical distance): in the design task to search alternative uses (Paper Clip, Water Bottle), it describes the designer's intended usage for the solution and the context it works in. Farther is the new context of use or more different is the solution, higher are the evidences of design space exploration, which also reflects into higher opportunities of using the idea for a patent application.
Distance (→ SAPPhIRE): whatever is the design task to face (problem- or solution- driven), it accounts for the degree of (dis)similarity among solutions. Its measurement requires a common reference to appreciate where are the differences. A significant difference between the generated idea and the reference concept reflects the skills of overcoming mental inertia and the attitude to generate ideas which go beyond the set of typically generated ones, which are not so novel or obvious for the majority.
Feasibility (→ technical plausibility): whatever the design task is, it classifies ideas depending of their chances to be developed, thus to appear sooner or later in the market.
Inventiveness (→ patent law): as for the last two mentioned, this also applies to every idea independently from the nature of the design task as all of them could be potentially suitable for a patent application.
The metrics has been subdivided into four sublevels as ordinal variables labelled for processing reasons with natural numbers (Table 1).

Table 1: The 4 metrics and related sublevels

Purpose	Distance	Feasibility	Inventiveness
1 - Same object	1 - Different Parts	1 - The solution seems against the laws of physics	1- obvious
2 - Same application field, different object	2 - Different Organ or Input	2 - The solution seems to need a technology yet to appear	2 - more obvious than inventive
3 - Same function, different application field	3 - Different Physical Effect	3- The solution requires some research activity in an established domain to get implemented	3 - more inventive than obvious
4 - Different function	4 - Different State change	4 - The solution just needs development or it's already available	4 - inventive

https://doi.org/10.1017/dsi.2019.57 Published online by Cambridge University Press

Four sublevels (or whatever even number of levels) trigger the students to express their opinion so that it is somehow flavoured, rather than neutral (which is possible with an uneven number of levels). Levels labelled 3 or 4 correspond to higher exploration and broader divergent thinking (Purpose, Distance) and higher chances to be patented (Feasibility and Inventiveness).

3 DATA GATHERING AND PROCESSING APPROACH

3.1 Gathering students' ideas - Design tasks

Since AY 2015/16 students of MTSI begin their practice lectures with several problems for which they are asked to provide solutions. Along these years, different design problems have been used to check the students' innate creative skills at the beginning of the class. The ideas collected during the first practice lecture, where students faced these design problems, represent the set to be considered for the ex-ante evaluation of students' inventive skills, so that it is possible to estimate their innate inclination to be creative before being exposed to the course contents.

Table 2: Distribution of design tasks/problems used in class with students in different AAYY

	2015/16	2016/17	2017/18
Peanut Sheller	X		
Coffee Capsule		X	X
Candle, matches and rings		X	X
Paper Clip	X		
Water Bottle		X	X
Patent (Wind Cups)			X

Table 2 provides details about when design problems have been used, along AAYY, to generate results for this study. The first three problems are characteristic of problem-driven design: ideas have to be generated to address a specific problem that satisfy a (more or less) clarified exigency. The remaining problems are examples of solution-driven design: ideas get generated in order to check where an already available solution can find a useful application in a different domain or context, to satisfy a not-yet-clarified exigency, that could have also remained latent so far. Beyond the co-evolutionary dynamics, if designing is about matching the problem space and the solution space (Dorst and Cross, 2001), problem-driven design can be seen as the attempt to match solutions to a known problem, while solution-driven design is the process of matching unknown problems to an existing solution. The following brief descriptions present the problems students faced.

Problem-driven/Peanut Sheller (Fu *et al*., 2010): "Design problem: Device to shell peanuts. In places like Haiti and certain West African countries, peanuts are a significant crop. Most peanut farmers shell their peanuts by hand, an inefficient and labor-intensive process. Our goal is to build a low-cost, easy to manufacture peanut sheller targeted at individuals and small cooperatives that will increase the productivity of the peanut farmers. Further, this peanut sheller should be manufacturable with materials that are readily available in the target communities. Our target throughput is approximately 50 kg/h. Customer needs: i) Must remove the shell with minimal damage to the peanuts; ii) Electrical outlets are not available as a power source; iii) A large quantity of peanuts must be quickly shelled; iv) Low cost; v) Easy to manufacture."

Problem-driven/Coffee Capsules (real industrial case study): Coffee capsules are for brewing high-quality espresso at home. More customers mean larger amount of exhaust capsules to dispose. Recycling is necessary. Coffee SMEs sell door-to-door, which facilitates the collection of exhaust capsules. Design a machine that recovers raw materials from the capsules. Requirements: i) separating the plastic shell in polypropylene, to be recycled, from the exhaust coffee and the thermowelded paper layers (both on top and bottom sides). Polypropylene has to be free of contaminants at the end of the operations; ii) preserving the exhaust coffee to sell it is a growing agent for the agricultural industry (no contamination, as for polypropylene) iii) processing 10 millions of capsules per year; iv) consuming a low amount of resources (energy and manpower).

Problem-driven/Two rings (adapted from McCaffrey and Krishnamurty, 2015): You are provided with a box full of matches, one candle and two rings. With these items you have to find solutions that allow you to link the two rings. The link has to be strong enough to withstand a 1kg load.

https://doi.org/10.1017/dsi.2019.57 Published online by Cambridge University Press

Solution-driven/Find alternative uses for this object: this is a common test to evaluate creativity. The task is: "find as many alternative uses for the object X as you can". Two objects have been proposed: paper clip and bottle of water.
Solution-driven/Patent as a source of inspiration is equivalent to alternative uses, as in this case the information does not come from an object as a whole, but from its documentation, where the solution is described together with its working principle. So far students have been asked to generate alternative uses for the solution described in patent US1833019 - "Airplane tire" (hollow projections on wheels' lateral surface allow them to start rotating before landing so as to reduce tread wear).
As it is clearly visible, problems changed year after year for reasons of renewal and extension of contents for the class. Nevertheless, the proposed problems are similar in nature (category by category). The coffee capsule problems can be considered virtually equivalent to the peanut sheller, as for both systems it needs to develop solutions capable of detaching two substances (peanuts and shells, the coffee and the capsule body) and sort the outcomes. Such a change is due to the larger availability of online information on the peanut sheller problem. The two rings problem should provide the students with an opportunity to think outside the box, e.g. by finding resources for the solution where they are not evidently available. Among the most used objects for the alternative uses task, it is worth to mention the brick and the paper clip. The increasing diffusion of the paper clip is the reason for its switch with the bottle of water. The extension of this test to patent-based data and information as patent structure and content analysis are part of the course syllabus and this allow students to familiarize with these documents.

3.2 Data processing

As this investigation has been carried out some year after the first collection of data, this allows to work on a reduced set of students to be considered as the reference cohort. This is to make a preliminary check of the suitability and the viability of the approach, also with reference to the relevance of emerging evidences, before repeating this on a larger scale.
To keep the analysis meaningful, the authors decided to focus on the students that got the best (approximately 10% per year) and the worst evaluations (approx. 10%/y) in the individual test, to compare against the outcomes of their idea generation activity at the beginning of the class. On average the class is composed by 50 students with small fluctuations in different AY.
For this preliminary study it is proposed to keep the analysis restrained to the last academic year (17/18), with the purpose of extending the analysis to previous years if the approach provides evidences of being effective and its results interesting to suggest direction for improvements to the teaching activity.
Two independently working coders carried out the assessment of the ideas with the metrics of Table 1. Before starting the coding process, they spent approximately two hours to clarify and share the meaning of the different classes and the sublevels therein. IRR is to be estimated by means of Fleiss' Kappa statistics (Fleiss, 1971) whose values of substantial agreement are above the score of 0.61. Scores above 0.81 reflect an almost perfect agreement (Landis and Koch, 1977). IRR values lower than 0.61 will require the coders to discuss about their different viewpoints and fix discrepancies by arbitration where possible or to run a second round of coding to fix discrepancies.
The purpose is to compare the ranking of students' ideas with their class ranking according to their final grade, under the assumption that a perfect positive correlation between the final grades and the initial conditions means that the teaching is not capable to improve the ideation and problem solving capabilities of less-talented students, as the students' with the best grades were also the one who resulted more inventive before being exposed to the course's contents. On the contrary a negative correlation would be an evidence of something opposite: the worst performing students at the beginning took so higher benefits of the contents from the syllabus that allow them to overcome their colleague which had a more brilliant start.
Students' ideas at the beginning of the class got collected by means of a form which triggers students to provide a comprehensive description of the idea, this is to provide the coder with elements to better interpret and evaluate the concept.
Moreover, it is worth to mention that some of the metrics described in Section 2 require the definition of a solution to be considered as a reference one, i.e. those which leverage comparisons between solutions in terms of mutual distance. Without this reference solution, the evaluation among coders can proceed randomly and not consistently. In order to define this reference solution, it was decided to scroll the

https://doi.org/10.1017/dsi.2019.57 Published online by Cambridge University Press

whole set of ideas and take as reference the most frequent one for that task. This is exclusively true for problem-driven design tasks. In solution-driven design tasks, as the "alternative uses", it is the object or the invention itself that works as reference for comparison.

Figure 2 clarifies the whole process of data gathering and processing. Figures 3 and 4 show two examples of forms filled in with students' ideas, respectively for the coffee capsules problem and the alternative uses for a bottle of water.

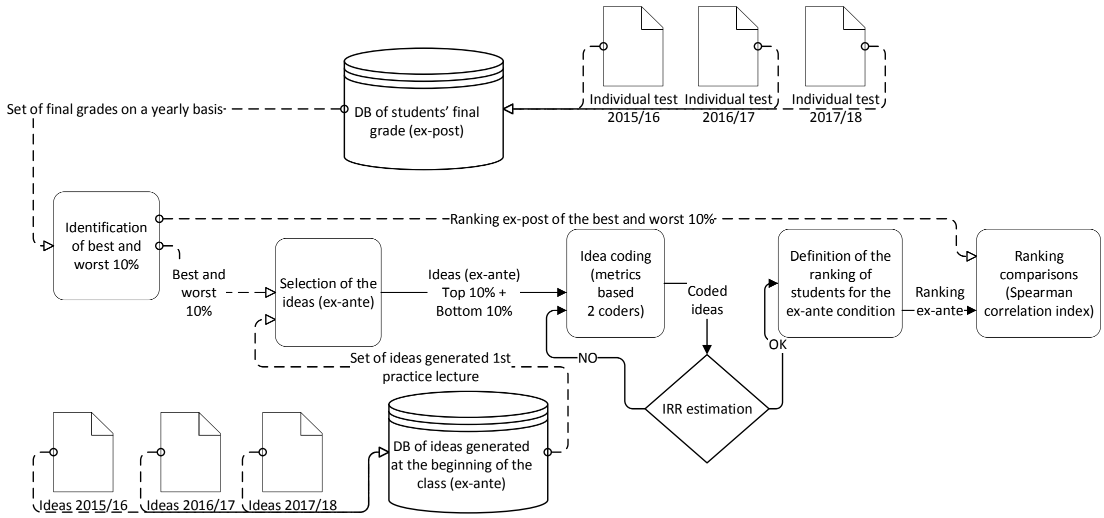
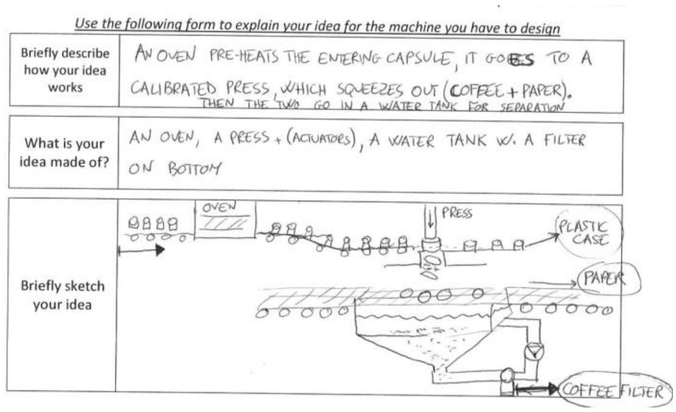


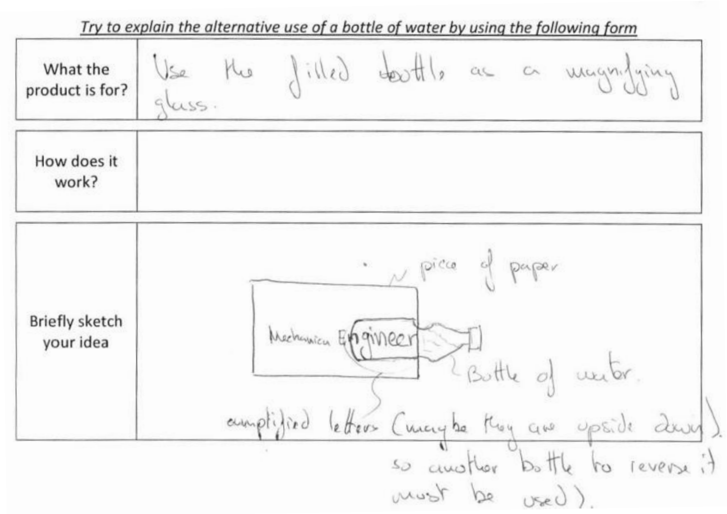
Figure 2: The overall process of data gathering and analysis



Use the following form to explain your idea for the machine you have to design

Briefly describe how your idea works	AN OVEN PRE-HEATS THE ENTERING CAPSULE, IT GOES TO A CALIBRATED PRESS, WHICH SQUEEZES OUT (COFFEE + PAPER). THEN THE TWO GO IN A WATER TANK FOR SEPARATION
What is your idea made of?	AN OVEN, A PRESS + (ACTUATORS), A WATER TANK W/ A FILTER ON BOTTOM
Briefly sketch your idea	OVEN; PRESS; PLASTIC CASE; PAPER; COFFEE FILTER

Figure 3: One among the generated ideas for the coffee capsules problem



Try to explain the alternative use of a bottle of water by using the following form

What the product is for?	Use the filled bottle as a magnifying glass.
How does it work?	
Briefly sketch your idea	piece of paper; Mechanical Engineer; Bottle of water; amplified letters (maybe they are upside down, so another bottle to reverse it must be used).

Figure 4: One among the alternative uses hypothesized for the bottle of water

https://doi.org/10.1017/dsi.2019.57 Published online by Cambridge University Press

4 THE DATA PROCESSED: RESULTS AND DISCUSSION

Figure 5 provides a detail view on the latest data processing steps presented in Figure 2. They describe how the ideas assessed with the metrics by the coders become a ranking of a cohort of students' performance before the class starts, which is the needed to run the comparison with students' final grades.

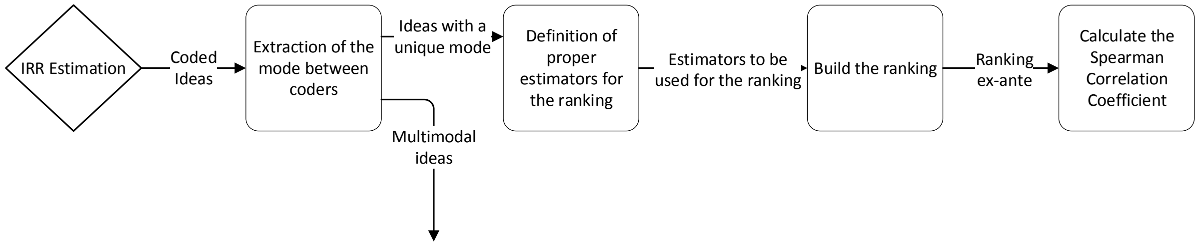


Figure 5: Details of the computational pipeline to create the ranking (ex-ante condition)

The initial alignment between coders, together with the definition of reference solutions to be considered for the ideas generated in problem-driven design tasks, allowed the coding process to be particularly effective and efficient, since Inter-Rater Reliability (IRR) scored 0,67 (Fleiss' kappa) considering the whole 4 sublevels for the different metrics, after the first round of coding. The value of IRR with the coders' evaluations polarized towards the extremes (1 and 2 → Low, 3 and 4 → High) raise to 85%. Nevertheless, the score 0,67 is to be considered acceptable according to Landis and Koch (1977). This means that on a total of 216 ideas generated by the cohort of twelve students, each idea was coded according to 4 different metrics with 4 sub-levels each. Each of the two coders made 864 decisions in front of an overall set of 3456 potential alternatives with a very satisfactory homogeneity. Such results of the coding stage is also an indirect evidence that the metrics is also effective to describe the ideas, as the stability of the coding scheme is typically measured this way. The two sets of coded ideas have been synthesized into a unique set of ideas that are characterized by having the same evaluation between coders and within the same metrics of evaluation. This means that ideas that coders evaluated not consistently within the same class have been neglected from the evaluation of scores concurring to the determination of that ranking (Purpose, Distance, Feasibility and Originality).

In order to find a reliable estimator which was capable of distinguishing the students performing better, it is also worth recalling that students were generally allowed to generate as many ideas as they wanted, as this is an evidence of good fluency (Kim, 2006). Therefore, this is also important to consider in order to better understand the individuals and their innate inventive/creative talent. However, the simple enumeration of the amount of ideas can be extreme biasing factor, as one brilliant and a bad idea would count the same. The sum of idea scores for each class of the metrics is definitely more precise in balancing the above limitation. It takes 4 ideas a very scarce interest to balance with a brilliant idea whose score is 4. This means that higher totals correspond to a better rank afterwards (higher position).

Table 3 presents several values. The first two columns from left present 12 students with anonymized names and results (final grades). The second column just show in which of the two tails of final grades distributions the selected students in the cohort were at the end of the class. The third column, instead, show the ranking of students final grades. Same numbers in the ranking correspond to the same final grades. For instance, subjects D, E and F have the same rank. This also means that there is no 8th, nor 9th place as 3 people share the 7th place, then the next free position is for the tenth (Subj C). The same rules have been applied to the sum of scores by category and by students (last four columns on the right). Therefore, for consistency reasons, also columns four to seven present ranking and not scores. The choice to shift from real numbers to ranking could be a biasing factor, as the ordinal nature of the ranking flattens differences (e.g. how far or how close two scores really are).

Row by row, the data in Table 3 also collects the rank for every student class of metrics by class of metrics. The last row, differently from previous ones, show the values of the Spearman's correlation coefficient for the four metrics. These coefficients of correlation have been calculated using the column of data where they are reported (e.g. as for the metrics of Section 2) and the ranking of the final grade displayed in column 3.

https://doi.org/10.1017/dsi.2019.57 Published online by Cambridge University Press

Table 3: Rankings and related correlation coefficients for the cohort of students of 2017/18

ID	Final Grade	Final Grade Rank (ex-post)	Purpose Rank (ex-ante)	Distance Rank (ex-ante)	Feasibility Rank (ex-ante)	Originality Rank (ex-ante)
Subj A	Lowest 10%	12	9	6	11	2
Subj B	Lowest 10%	11	12	7	12	12
Subj C	Lowest 10%	10	9	12	9	6
Subj D	Lowest 10%	7	6	7	9	2
Subj E	Lowest 10%	7	3	5	4	6
Subj F	Lowest 10%	7	7	7	6	6
Subj G	Highest 10%	5	4	3	5	6
Subj H	Highest 10%	5	9	4	3	5
Subj I	Highest 10%	3	1	1	2	1
Subj J	Highest 10%	3	2	2	1	6
Subj K	Highest 10%	2	5	7	6	2
Subj L	Highest 10%	1	8	11	6	6
Spearman's correlation coefficient			**0.59**	**0.25**	**0.76**	**0.30**

Values close to zero for the Spearman's correlation coefficient are probably the best ones, as they should show that the students substantially harmonized their skills: those having more troubles in generating convincing solution concepts at the beginning of the class should be now capable of achieving results as they were born inventive.

In detail, for what concerns what measured with metrics, the results of the correlation coefficient analysis triggered some interesting insights.

The Spearman's coefficient for the ranking based on the "purpose" metrics is an evidence that the situation at the end of the class has just slightly changed. More talented students are the ones more capable to look for solutions in other contexts, but there is a partial inversion in the trend, as this score is approximately 0,6. This results also highlights that there is further room to leverage this skill in students (that it is worth recalling it gets only measured for solution-driven design tasks), which opens up different strategies to improve the effectiveness of the teaching (e.g. by increasing hours devoted to the practical activities using specific tools and so forth).

"Distance" potentially accounts for the capability of students to overcome their mental inertia and explore alternatives at various detail level. This low value (0.25) reflects the effectiveness of teaching as this initial ranking is poorly correlated to the final one. This is a very satisfactory outcome, as many topics within the syllabus aim at improving the students' skills that enable them to explore the design space in the search for alternatives more effectively and more efficiently, so that they quickly overcome the stage where they generate typical solutions and shift generating non-obvious ones (also because of their improved capabilities of abstraction and analysis).

The relatively high value for metrics focusing on feasibility is, however, the highest among the four. It scores 0.76, which reflects a still strong direct correlation between the ex-ante and the ex-post condition. Among the 4 considered metrics, however, this is the less surprising result, as it is expected that students producing the solutions which are implementable or somehow attainable in the short term at the beginning are also the ones having a more solid background and common sense built along their learning path, in previous class and with previous experiences.

The correlation for the metrics about inventiveness, similarly to distance, has a quite low value (0.3) which shows that the course provided an effective contribution towards inventiveness. This result is a preliminary evidence that also those who initially showed to be less talented in problem solving in design and idea generation activities, at the end of the class manage to reach satisfactory results. which is a further evidence of an improved skill of exploration at the end of the course, looking for what is typically considered "non-obvious".

5 CONCLUSION

The paper presents an approach to measure the impact of design education in courses dealing with open ended problems, whose solutions are typically influenced by individual talent. The approach is tested on a small cohort of students from a class on Methods and Tools for Systematic Innovation at

https://doi.org/10.1017/dsi.2019.57 Published online by Cambridge University Press

Politecnico di Milano, a course attended the Advanced Mechanical Design track within the MS in Mechanical Engineering. The approach proposed by the authors requires measuring the inventive skills of students at the beginning of the class, before they get exposed to design methods and tools to stimulate divergent and convergent thinking. This measurement is to be compared with the final grades of students, in order to identify correlations between the two conditions (ex-ante/ex-post). Higher values of direct correlation correspond to lower impact of the teaching activity and vice versa. The measurement of ideas is carried out through tailored metrics that adapt idea evaluation metrics from the design creativity field and mixes them with relevant concepts capable of describing the impact an idea of solution might have in the industrial arena.

The approach proved to be effective to identify the of areas of teaching (i.e. by teaching objective) where the effectiveness of learning made better or worse performance, which reflects a more or less significant impact of the contents presented during the class. Within this perspective, this approach allows teachers, professors and other profiles who play a role in the education process to keep the effectiveness of the teaching monitored. This also makes possible to identify directions, objectives or topics which should deserve changes in the teaching routine or in the way the lessons are organized.

This promising results suggest that the study is worth of extension to the whole class for the three academic years for which data have been collected and made available. Nevertheless, the process of making the assessment of students' ideas for creating the ranking about the ex-ante condition is particularly time-consuming and labour intensive, which suggest for the possibility to train a higher number of coders to parallelize the operations and get results quicker (even within the academic year).

As briefly mentioned above, the approach is repeatable in different classes dealing with design methods and tools and, more in general with open ended problems whose solution also reflects individual talent.

As this study require a metrics tailored on the class objectives, the adaptation to a different syllabus would require an adaptation of the metrics to run the ex-ante evaluation of ideas.

REFERENCES

Becattini, N., Borgianni, Y., Cascini, G. and Rotini, F. (2012), "Model and algorithm for computer-aided inventive problem analysis". *Computer-Aided Design*, Vol. 44 No. 10, pp. 961–986.

Becattini, N. and Cascini, G. (2016), "Improving self-efficacy in solving inventive problems with TRIZ". In *Multidisciplinary Contributions to the Science of Creative Thinking*, (pp. 195–213). Springer, Singapore.

Becattini, N., Cascini, G., O'Hare, J.A. and Masclet, C. (2018), "Coding schemes for the analysis of ICT supported co-creative design sessions". In *DS92: Proceedings of the DESIGN 2018 15th International Design Conference*, (pp. 533–544).

Dorst, K. and Cross, N. (2001), "Creativity in the design process: co-evolution of problem–solution". *Design studies*, Vol. 22 No. 5, pp. 425–437.

Fleiss, J.L. (1971), "Measuring nominal scale agreement among many raters". *Psychological bulletin*, Vol. 76 No. 5, p. 378.

Fu, K., Cagan, J. and Kotovsky, K. (2010), "Design team convergence: the influence of example solution quality". *Journal of Mechanical Design*, Vol. 132 No. 11, p. 111005.

Gero, J.S. and Kannengiesser, U. (2004), "The situated function–behaviour–structure framework". *Design studies*, Vol. 25 No. 4, pp. 373–391.

Jia, L., Becattini, N., Cascini, G. and Tan, R. (2018), "Testing design stimuli for design-byanalogy on a large set of designers". In *DS 89: Proceedings of The Fifth International Conference on Design Creativity* (*ICDC* 2018), University of Bath, Bath, UK, (pp. 39–46).

Kim, K.H. (2006), "Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT)". *Creativity research journal*, Vol. 18 No. 1, pp. 3–14.

Landis, J.R. and Koch, G.G. (1977), "The measurement of observer agreement for categorical data". *Biometrics*, pp. 159–174.

McCaffrey, T. and Krishnamurty, S. (2015), "The obscure features hypothesis in design innovation". *International Journal of Design Creativity and Innovation*, Vol. 3 No. 1, pp. 1–28.

Saliminamin, S., Becattini, N. and Cascini, G. (2018), "Sources of creativity stimulation for designing the next generation of technical systems: correlations with R&D designers' performance". *Research in Engineering Design*, pp. 1–21.

Shah, J.J., Smith, S.M. and Vargas-Hernandez, N. (2003), "Metrics for measuring ideation effectiveness". *Design studies*, Vol. 24 No. 2, pp. 111–134.

Srinivasan, V. and Chakrabarti, A. (2010), "Investigating novelty–outcome relationships in engineering design". *AI EDAM*, Vol. 24 No. 2, pp. 161–178.