Diverse Cognitive Skills and Team Performance

A Field Experiment based on an Entrepreneurship Education Program

Huber, Laura Rosendahl; Sloof, Randolph; Van Praag, Mirjam; Parker, Simon C.

Document Version
Accepted author manuscript

Published in:
Journal of Economic Behavior & Organization

DOI:
10.1016/j.jebo.2020.06.030

Publication date:
2020

Citation for published version (APA):
Huber, L. R., Sloof, R., Van Praag, M., & Parker, S. C. (2020). Diverse Cognitive Skills and Team Performance: A Field Experiment based on an Entrepreneurship Education Program. *Journal of Economic Behavior & Organization*, *177*, 569-588. https://doi.org/10.1016/j.jebo.2020.06.030

Diverse Cognitive Skills and Team Performance: A Field Experiment Based on an Entrepreneurship Education Program*

Laura Rosendahl Huber[a,b] (corresponding author)

Randolph Sloof[c,d]

Mirjam Van Praag[e,d,f]

Simon C. Parker[g,f]

[a] Rotterdam School of Management, Erasmus University, Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands. Email: rosendahlhuber@rsm.nl
[b] Max Planck Institute for Innovation and Competition, Marstallplatz 1, 80539 München, Germany.
[c] Amsterdam School of Economics, University of Amsterdam, Roeterstraat 11, 1018 WB Amsterdam, The Netherlands. Email: r.sloof@uva.nl
[d] Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam, The Netherlands
[e] Copenhagen Business School, Dept of Innovation and Organizational Economics, Kilevej 14a, 2000 Frederiksberg, Copenhagen, Denmark. Email: mvp.si@cbs.dk
[f] IZA Institute of Labor Economics, Schaumburg-Lippe-Strasse 5-9, 53113 Bonn, Germany
[g] Ivey Business School, Western University, 1255 Western Road, London, Ontario, Canada. Email: sparker@ivey.ca

*We would like to thank Thorsten Grohsjean and Karin Hoisl for very useful comments and suggestions. We also thank seminar audiences at the 5th IZA Workshop on Entrepreneurship research (Potsdam), SEI Doctoral consortium (Milan), SMS conference (Tel Aviv), EALE conference (Turin), Amsterdam Center for Entrepreneurship (University of Amsterdam and VU University) and Tinbergen Institute for helpful comments.

DIVERSE COGNITIVE SKILLS AND TEAM PERFORMANCE:
A FIELD EXPERIMENT BASED ON AN ENTREPRENEURSHIP EDUCATION PROGRAM

Abstract

Verbal and mathematical reasoning are key cognitive skills which individuals use throughout their lives to create economic value. We argue that individuals undertaking entrepreneurial tasks also draw on these skills, and we study how best these skills should be combined in entrepreneurial teams. To that purpose we conduct a randomized field experiment using data from the BizWorld entrepreneurship education program. Four different types of teams are created which differ in terms of their cognitive skill composition. Our results show that balanced skills are beneficial for a team's venture performance only if it comes from within-person skill balance, and that combining team members with different skills in mixed teams does not compensate for a lack of members who individually possess balanced cognitive skills.

Keywords: skill balance, entrepreneurship, team performance, team diversity, field experiment
JEL-codes: J24, L25, L26, M13

1 Introduction

Scholars are paying growing attention to the composition of entrepreneurial teams, reflecting the prevalence of such teams among new venture founding efforts (Kamm et al., 1990; Brannon et al., 2013; Klotz et al., 2014). Teams may be able to share and exploit valuable complementary resources, such as skills, social networks and finance (Kim and Longest, 2014; Jin et al., 2017). Indeed, evidence suggests that entrepreneurial teams may enjoy superior venture performance relative to solo founding efforts (Kroll et al., 2007; Åstebro and Serrano, 2015). Moreover, there is increased awareness of the importance of founding team member characteristics for strategic decision making and venture performance (Ucbasaran et al., 2003; Zheng et al., 2016; Jin et al., 2017; Kollmann et al., 2017).

In prior research, diversity of team member attributes ('heterophily') has been linked to bridging social capital and venture performance advantages (Kim et al., 2005; Parker, 2009; Steffens et al., 2012; Eesley et al., 2014). The attributes focused upon in these studies are typically socio-demographic in nature (e.g. Chowdhury, 2005; Horwitz and Horwitz, 2007; Foo, 2011). Also within the broader team diversity literature, the impact of demographic diversity on team performance has often been studied. The primary reason for doing so is derived from the *informational diversity – cognitive resource perspective*, i.e., the presumption that demographic diversity implies diversity in knowledge, skills and perspectives, and thus increases the overall amount of (valuable) cognitive resources the team can draw on (cf. Webber and Donahue, 2001; Bell et al., 2011). Taking this perspective, cognitive skills - and human capital more broadly - constitute the more fundamental attribute to assess the impact of diversity. Existing entrepreneurship studies have demonstrated the importance of human capital, and cognitive skills in particular, for performance on entrepreneurial tasks at the individual level (Wright et al., 2007; Van der Sluis et al., 2008; Hartog et al., 2010). Moreover, diversity among team members in this key determinant of entrepreneurial performance also seems to enhance team performance, although the evidence is scarce and mixed (Vanaelst et al., 2006; Hoogendoorn et al., 2017). A question that has not yet been studied, however, is the importance of individual and team level *skill balance* for entrepreneurial team performance.

At the same time as research on entrepreneurial teams has been gathering pace, a largely parallel stream of scholarship has been exploring the impact of entrepreneurship education programs on entrepreneurial intentions and founding efforts (Von Graevenitz et al., 2010; Martin et al., 2013). These programs are usually conducted in team settings, such as school-age Junior Achievement programs (Oosterbeek et al., 2010; Elert et al., 2015) and university-level entrepreneurship courses (e.g. Premand et al., 2016; Hoogendoorn et al., 2017). Yet although team composition issues may play an important role in the success of these programs, to date few researchers have studied them. Moreover, the impacts of these programs have usually been evaluated at the individual level, rather than the team level at which these programs are usually implemented. Hence, despite the importance of both team entrepreneurship and entrepreneurship education in the academic literature as well as business practice, we currently lack evidence about how important aspects of team diversity are related to the performance of teams formed as an integral part of those programs (Foo, 2011).

Diversity can have many dimensions; in line with the cognitive resource perspective, the focus of this article is on cognitive skill diversity. In the entrepreneurial setting, theories about how skills affect entrepreneurship have been framed at the individual level (Marvel et al., 2016; Lazear, 2005), and more recently also at the team level, reflecting the prevalence and importance of entrepreneurial teams (Klotz et al., 2014). To gain a better understanding of the optimal composition of entrepreneurial teams, this article explores the effect of skill composition on the performance of teams undertaking an entrepreneurial task, focusing on a specific type of cognitive skill diversity: skill balance, both within and across team members. Several studies have highlighted the importance of balanced skills for entrepreneurship at the individual level (Lazear, 2005; Chen and Thompson, 2016; Åstebro and Yong, 2016; Aldén et al., 2017; Patel and Ganzach, 2019). However, the impact of skill diversity at the team level is much less clear cut (Kaiser and Müller, 2015; Jin et al., 2017; Kollmann et al., 2017). This may in part be due to earlier studies using biased measures of skills and skill diversity, as these are often based on observational data. That is, individual skill levels are typically proxied with education and experience related measures (Marvel et al., 2016). The problem with such (investment based) measures is their endogenous nature, because experience and educational trajectories involve choices. Similarly, demographic diversity has often been used as indirect measure of skill diversity (Webber and Donahue, 2001; Bell et al., 2011). The widely scattered results within the team diversity literature might partly be due to this indirect operationalization, and thereby "oversimplification", of the true underlying diversity concept these studies aim to capture (cf. Bell et al., 2011, p. 730).

In this paper we aim to study the impact of cognitive skill diversity using unbiased measures. In particular, we address the following two research questions: 1) Do teams with balanced cognitive skills perform better than teams with unbalanced cognitive skills?

2) Given that individuals can have balanced or unbalanced skills, is the source of skill balance at the team level important, i.e., is skill balance inherent in each team member as effective as combining different skills from team members with different unbalanced skills?

We explore these questions in the context of a well-known international school-age entrepreneurship education program called BizWorld. This particular setting allows us to tackle two key challenges which beset empirical research on team entrepreneurship: selection bias and survival bias. First, selection bias may arise because individuals often choose whether to join teams and which teams to join; and team composition may be endogenous to performance outcomes (Foo, 2011). These challenges may confound causal inference from survey questionnaires. We sidestep these challenges by conducting a field experiment in which team membership is randomized exogenously and participants are similar with regard to salient attributes. This field experiment therefore adds to the limited number of studies which have systematically explored the performance impacts of exogenous random variation of entrepreneurship education teams (Premand et al., 2016; Hoogendoorn et al., 2017). Second, by using teams in an educational setting, our results are not affected by survival bias due to high failure rates among early start-ups that might be related to the characteristics of the founding team (Jin et al., 2017).

Two further advantages of the design of our field experiment may be noted. First, in contrast to prior field studies, ours enables the direct and objective measurement of human capital outcomes (i.e., verbal and mathematical skills) with a high degree of accuracy. This is beneficial because investment-based human capital indicators (such as years of education or work experience) are viewed as indirect predictors of human capital and therefore less predictive of the entrepreneurship-success relationship (Unger et al., 2011; Marvel et al., 2016). Moreover, subjective assessment of individual ability and expertise could be a confounding factor when studying team performance. Second, the field experiment enables us to identify causal impacts cleanly since youths lack heterogeneous labor market experience and have to rely on their cognitive skills. Hence, we can largely bypass problems caused by the endogeneity of experience and preferences. Of course, as is invariably the case with field experiments, the specialized nature of this experimental setting may come at the price of limited external validity; we go on to discuss this point in detail towards the end of the paper.

The results from our field experiment provide some evidence that having diverse (i.e. balanced) cognitive skills within the team fosters entrepreneurial team performance. Our most important finding, however, is that in a team setting it crucially matters how skill balance is achieved; teams perform better only if skill balance results from having team members with individually balanced skills. A tentative explanation - based on a more exploratory analysis - is that also within entrepreneurial teams there likely exist distinctive roles or tasks that benefit from having individually balanced skills. Having some jack-of-all-trades within the team's ranks then facilitates effective task allocation.

The remainder of the paper has the following structure. We start by developing hypotheses about how intra-team diversity in two salient forms of cognitive skills, namely math and verbal, influences entrepreneurial team performance. The BizWorld program is then described, followed by an account of the field experiment and methodology. The results are then presented, before the paper closes with a concluding discussion in which we summarize our findings and discuss the limitations of our study.

2 Theory and Hypotheses

We start by defining two fundamental cognitive skills at the individual level: mathematical and verbal skills. We draw on entrepreneurship research to briefly discuss how these skills have been linked to entrepreneurial performance at the individual level. We then go on to relate the diversity of cognitive skills to entrepreneurial team performance.

2.1 Cognitive skills and entrepreneurial performance at the individual level

Several studies in the entrepreneurship literature have shown the importance of mathematical skills for entrepreneurial performance. For example, Hartog et al. (2010) estimate that higher returns in entrepreneurship relative to paid employment are most strongly related to mathematical and technical ability. Moreover, Almus and Nerlinger (1999) estimated that founders with degrees in technological subjects generated higher employment growth than those with degrees in 'softer' subjects; while Van Praag and Cramer (2001) estimated that science-oriented education was positively related to firm size. Finally, Mueller (2006) estimated that fields of study like engineering and natural sciences were positively associated with R&D, which in turn is often found to be a predictor of innovation and economic growth (Aghion and Howitt, 2006).

In addition, several other studies have highlighted the importance of verbal ability for business outcomes. One of the most salient applications of verbal ability is through the use of language in communications. For instance, the seminal paper by Aldrich and Fiol (1994) describes how the type of language used by an entrepreneur and the way he or she communicates about his/her new business, product, or technology influences the venture legitimization process and the socio-political approval for the entrepreneur's activities. Especially when entering brand new markets and when introducing innovations, entrepreneurs need to build a comprehensive story about the potential of their innovations and how they will benefit stakeholders (Howell and Higgins, 1990). Additionally, one of the most crucial determinants of venture performance is the extent to which an entrepreneur is able to secure financing. Research has shown that entrepreneurs need verbal skills to communicate effectively with resource-providers and that spinning narratives helps them to obtain financing (Martens et al., 2007; Parhankangas and Ehrlich, 2014; Allison et al., 2015). Finally, language and communication (both in written and in oral form) are also essential for marketing to engage with customers and other important stakeholders (Lounsbury and Glynn, 2001; Gao et al., 2016).

Besides studying the effect of specific skills in isolation, the optimal skill combination for entrepreneurs has also received considerable attention at the individual level. Following the influential contribution of Lazear (2005), several papers have studied the effect of skill balance, or being a Jack-of-All-Trades (JAT), on both the choice to become an entrepreneur (e.g., Wagner, 2006; Chen and Thompson, 2016) and on entrepreneurial performance (e.g., Hartog et al., 2010; Åstebro and Thompson, 2011; Åstebro and Yong, 2016). The main idea behind Lazear's JAT theory is that, because entrepreneurs engage in a wide variety of tasks, they require (at least basic) knowledge of numerous business areas (Aldén et al., 2017; Kacperczyk and Younkin, 2017). For example, when founding a business not only the technical conception and development of an idea is required, but also an entrepreneur must be able to analyze trends, communicate with employees, and convince customers and other stakeholders of the value of the idea (Strohmeyer et al., 2017). Furthermore, due to the rapidly changing environment and uncertainty surrounding new venture creation, it is unclear ex-ante which specific skill would be most beneficial (Baron and Henry, 2010). Thus, having a diverse set of skills will enable the entrepreneur to deal with the variety of challenges that she is likely to face, without having to rely on external resources of an often, ex ante, unknown type. In this view, employees can be specialists, but entrepreneurs benefit from having a balanced set of skills (Lazear, 2005). Moreover, even though entrepreneurs can hire other people, the individual entrepreneur must still possess a variety of skills to be able to judge other people's abilities and how to combine them. In sum, "...entrepreneurs need not excel in any one skill but [should be] competent in many" Lazear (2005, p. 649).

Lazear's JAT theory has been tested empirically in terms of entrepreneurship entry and performance. Positive results on entry have been found by Lazear (2005), Wagner (2006) and Lechmann and Schnabel (2014). Several recent studies have also found a positive relationship between individual balanced skills and entrepreneurial performance (e.g. Hartog et al., 2010; Åstebro and Yong, 2016; Aldén et al., 2017; Patel and Ganzach, 2019). Diversity in cognitive skills thus indeed seems to pay off for solo entrepreneurs.[1]

[1]The importance of having a broad skill set (or being a *generalist*) has also been shown for high level managerial roles (see e.g., Ferreira and Sah, 2012; Custódio et al., 2013). Since both entrepreneurs and managers are responsible for strategic and complex decisions with a high level of unpredictability, this result seems to be consistent with the results obtained in the entrepreneurship literature.

2.2 Diversity in cognitive skills and entrepreneurial team performance

As noted above, prior research on balanced skills (Lazear, 2005) has been framed at the individual level. This theory cannot be simply translated to the team level, because teams comprised of two or more individuals possess different skill sets across, as well as within, individuals. Hence cross-person skill composition must be considered in addition to within-person skill balance. For example, a deficiency in one type of skill by one team member may be compensated by another team member who possesses that skill. More generally, in the team entrepreneurship context, the relevant unit of analysis is the venture, not the individual (Ucbasaran et al., 2003). Nevertheless, note that the JAT logic for solo entrepreneurs essentially derives from entrepreneurial tasks being characterized by diminishing returns to particular (cognitive) skills. The latter argument can be equally well applied to team ventures.

Teams in a start-up setting are required to perform a wide range of different tasks, which call on different skill sets to execute. In terms of the field experiment described in this paper, a mix of verbal and mathematical skills are likely needed to determine key parameters of the businesses and generate sales. Thus, mathematical and logical skills are needed to keep accounts, calculate costs and prices, and plan the structure of the business and its operations. At the same time, verbal skills are needed to design the marketing plan and persuade customers to buy the venture's product. Similar to arguments articulated by Kremer (1993), whose 'O-ring' theory stressed the vulnerability of entire composite production processes to apparently minor component failures, the success of business ventures can only be as good as their weakest link. For example, an unbalanced team where verbal skills are pronounced but math skills are in short supply may sell lots of products; but if they lack the math skills to cover their costs per unit, they will only end up making a greater loss. Here, the weakest link is math skills. Conversely, an unbalanced team where math skills are abundant but verbal skills are in short supply may set a price that delivers a healthy sales margin; but if they lack the verbal skills to sell much product, they will generate low sales volumes. Here, the weakest link is verbal skills. In contrast, a team possessing a balance of math and verbal skills can discharge both tasks satisfactorily, yielding performance benefits relative to the other two cases.

Skill balance at the team level will be more likely to promote superior team performance if there are diminishing returns to particular skills. In particular, balanced skills maximize team productivity when the total stock of skills available to the team are in fixed supply (as in our field experiment), and when team returns to skills are increasing and concave functions of those skills. Diminishing returns in the form of concave production functions are known to characterize virtually all real-world production processes (Shephard, 2012).[2] Based on these arguments we obtain:

Hypothesis 1 *Conditional on having equal average ability, teams with a balance of cognitive skills enjoy superior performance compared with teams lacking a balance of cognitive skills.*

Given the importance of skill diversity for entrepreneurial team performance as described above, another important question concerns the source of the skill balance at the team level. Consider for example two types of teams with the exact same skill set when aggregated over team members. One of the teams comprises several members *each of which* has individually balanced skills, whereas the other combines members with different types of individually unbalanced skills. Call these teams 'Type IB' and 'Type IU', respectively. In the light of Hypothesis 1, the second kind of team might be constructed by team founders as a logical response to a lack of skill balance within individuals. One can then ask: Are these teams likely to perform as well as each other?

Being formulated at the individual level, Lazear's JAT theory does not directly address this question. Nevertheless, the broader JAT logic that entrepreneurial ventures require various cognitive skills to be *effectively* combined and work together can arguably also be applied here in a similar way. A first, direct way to do so is fully in the spirit of Lazear's theory. Even though in team startups members are typically equal and task

[2] Of course, it needs to be recognized that skill balance is only one driver of team performance: another is the mean overall ability of a team. Mean ability within a team thus should be controlled for when isolating the effects of skill balance, and we do so accordingly in our experiment and empirical analysis (cf. Sections 4 and 5).

are highly interdependent, with all members to various degrees involved in all tasks, members will nevertheless focus to some extent on particular tasks and perform different roles (if only so informally). Some of these roles may require balanced skills per se. This occurs, for instance, if one of the team members (informally) takes up the role of coordinating the various tasks among the different members within the team. In line with the JAT theory at the individual level, this person needs a diverse set of skills in order to judge and combine the skills within the team most effectively. Alternatively, there might be tasks within the team that require a certain level of all relevant skills in order to perform this particular task optimally. As an example, in order to be able to understand the financial decisions that have to be made for the business to be successful, as well as being well-versed enough to communicate and explain these decisions to the other team members, a sufficient level of both math and verbal skills might be required. By their very composition, IB teams have individuals available that could well perform such tasks requiring diverse (i.e. balanced) skills, whereas IU teams do not. If there is value in having balanced skills at the level of the entrepreneurial venture (cf. Hypothesis 1), IB teams have the additional advantage that balancedness is also embodied in their individual members, which facilitates effective task allocation. This suggests that IB teams are likely to perform better than IU teams do. Therefore we propose:

Hypothesis 2 *Teams where skill balance is derived from combining individuals who individually possess balanced skills, enjoy superior performance compared with teams where skill balance is derived from combining individuals who individually possess unbalanced skills.*

3 Program and context

The subjects of our field experiment are teams of school pupils participating in one of the leading entrepreneurship education programs in the world, called the BizWorld program. In the Netherlands the program started in 2002 and approximately 40,000 children had participated in it at the time of the experiment, in the spring of the school year 2009-2010;

these numbers have grown exponentially since then. The sample used for our experiment included participating schools in the proximity of Amsterdam where the experiment was conducted, which enabled us to monitor the schools closely.

The Dutch version of the program is taught by an entrepreneur (or someone from the business world) in cooperation with the class teacher. The BizWorld Foundation matches schools and sponsoring entrepreneurs willing to participate. Thus, financial or network constraints do not hinder schools' voluntary participation in the program. The minimum level of participation is an entire class, i.e., individual pupils or teams cannot participate. Prior to the start of the education program the teacher and the entrepreneur are introduced to each other during a train-the-trainer session. During this session the content of the program is explained and the teacher and the entrepreneur receive the "BizWorld-suitcase", containing a detailed course handbook and course materials. Parents were also informed about our research project and field experiment at this time.

The pupils participating in the program are aged 11 or 12. The five-day program usually takes place within a period of 2 to 4 weeks during the last few months of the school year. It is a structured program and different skills are crucial for performance, as the description of the program will demonstrate. A day-by-day overview of the content is shown in Table 1. Further details can be found at the Bizworld website: www.bizworld.org.

[INSERT TABLE 1 AROUND HERE]

Teams of 5 or 6 pupils are formed on the first day of the program, where each team serves as one company. Usually, teachers determine the team composition; but in this field experiment we were allowed to compose the teams, as explained in Section 4.2. Each team member is assigned a specific role within the team based on his or her preferences and the assessment of the teacher. The different roles to be fulfilled are: General Manager (CEO), Finance Director (CFO), Director of Product Design, Director of Production, Marketing Director, and Sales Director. Even though each of these roles comes with their own loosely described tasks and responsibilities (e.g., the marketing director is responsible for the image of the company and the marketing campaign, and the CFO is responsible for keeping track of the finances on a daily basis), all team members are to various degrees involved in all company tasks and decisions. In particular, while the whole team is involved in the design, production, marketing and sales of the products, only the CEO and the CFO present the business plan to the venture capitalist and decide how many shares to sell. The strong interdependence of tasks makes that collaboration and team work are crucial during the entire program.

On the second day, all the teams write a business plan, which is presented to a "venture capitalist" (i.e., teacher) in order to sell stocks and raise start-up capital. The quality of the business plan together with the presentation determine the share price the investor is willing to pay. The evaluation of the teams determining share prices is done by means of a very structured scheme that contains several criteria (i.e., business and employee engagement, presentation quality, and product and sales strategy). The use of this scheme (and the use of the teacher spreadsheet as explained in Section 4.2) leave little scope for subjective and/or biased measurements of team performance by the teachers. Teams also prepare the design and a prototype of their product on the second day.

It is important that everything is well planned and prepared for the production process which is scheduled for the third day. During the production phase the teams have a limited amount of time (one hour) to manufacture as many products as possible. When the production process is over, the total production costs are calculated. On the basis of these costs, the teams have to determine the sale prices of their products. From the third day onward, the teams have two alternative routes to raise more capital. They can either sell more shares to the venture capitalist, thereby reducing their ownership share in the company, or they can take up a loan from the bank which has to be redeemed, including interest, before the end of the program.

On the fourth day teams design their marketing campaign, which consists of a poster, a slogan and a "commercial". The commercial can either be a short stage play or a short video. Each team is given the opportunity to present their "commercial" in front of the group of prospective buyers before the “big sale” starts. The big sale is an organized fair at which the products are sold to the children in the grade below. After the sales market is over, the revenues are calculated.

On the last day of the program, each team has to complete a financial report consisting of a profit and loss statement and a balance sheet. At the end of this day the winning team is announced and rewarded. Individual team members have strong incentives to care about the business performance of their team. First, the BizWorld foundation provides certificates for each member of the winning team. Second, the entrepreneur or the company that sponsors the education program at the school offers some small prizes (usually in the form of gadgets) to the winning team. Third, but only in the year we conducted the field experiment, team members of the winning team were awarded a gift voucher of 7.50 euros each, and the members of the runner-up team were each awarded a gift voucher of 5.00 euros. These incentives are quite significant given that the average amount of pocket money for children at that age is 4 euros per week or 15 euros per month.[3] Finally, based on our observations, BizWorld teams usually show a strong motivation to achieve good company results and win the competition in their class.

As can be seen from the brief overview of the education program, different skills are required for the successful execution of different elements of the program which together determine team performance. For example, verbal skills are important for successful written and oral communications, such as writing and presenting the business plan as well as the design and execution of the sales campaign, whereas mathematical skills are important for planning the production process, financial reporting and financial decision making.

4 Data and Methodology

Our field experiment exogenously composed four distinct types of teams using a two-step procedure. Before turning to specifics, we first explain the general approach. All children were first classified as being one of three individual types: children with individual balanced skills (IB), children with individually unbalanced skills having better math than verbal skills (IU-math), and children having better verbal than math skills (IU-verbal).

[3]This information was collected for the children in our sample through a questionnaire before the start of the program.

In a second step, different combinations of these individual types were sorted into four different types of teams: individually balanced (IB) teams, individually unbalanced math (IU-M) teams, individually unbalanced verbal (IU-V) teams, and individually unbalanced mixed (IU) teams. Figure 1 illustrates this, assuming for ease of exposition that there are only three possible levels for each skill: high (H), medium (M), and low (L). IB individuals are the ones having equal math and verbal skills. Nevertheless, there is still considerable variation among IB individuals, as some of them are very good in both subjects (the IB_H types in the first line of Figure 1) whereas others are good in neither of them (IB_L). A similar (within group) variation applies to the IU individuals. These are the ones who have a *comparative* advantage in one of the subjects. Yet this does not necessarily imply that they are all very good in their best subject, i.e., they need not have an *absolute* advantage. In Figure 1, for instance, the Math_{ML} individuals are worse in mathematics than e.g., the IB_H individuals and equally good as Verb_{MH}.

[INSERT FIGURE 1 AROUND HERE]

Team composition subsequently took place at the class level. To ensure comparability, we aimed at keeping teams within one class as comparable as possible in other dimensions, most notably in terms of average overall skills. We illustrate this by again using Figure 1 (left column). With the top, middle and bottom panel hypothetically representing three different teams with 3 members each, the three teams have equal average overall skills (two times level H, two times level M and two times level L). Unavoidably, by design the IU-math teams have higher average math skills (the average of H, H and M), but at the same time lower average verbal skills (M, L and L). The flip side holds for the IU-V teams. IU teams in this illustrative example would be composed of either (Math_{ML}, Math_{HM}, Verb_{LH}) or (Verb_{LM}, Verb_{MH}, Math_{HL}), thus having equal average skills in both dimensions as the IB teams. Because general ability has been shown to matter both for individual and team entrepreneurial performance (Hartog et al., 2010; Foo, 2011; Jin et al., 2017; Patel and Ganzach, 2019), we designed our experiment in such a way that the average overall skills of the different teams that competed within a class was (as close as possible to being) constant, while at the same time forming as many "usable" teams as possible.

4.1 Cognitive skills and individual types

To measure cognitive skills, we used information about individual pupils' verbal and mathematical skills either from a national exam or a standardized student assessment system. The skills measured in these assessments are similar to the ones defined by the Common Core Standards used by BizWorld and related to the tasks in the program. The purpose of the nationwide exam and the standardized assessment system is to provide an objective assessment of the academic capabilities (and potential) of the pupils. That is, they are designed to measure the skills that an individual pupil has acquired during the years he/she was in primary school. Moreover, the performance on these standardized tests also provides an indirect indication of other important performance related traits such as intelligence, attention control, motivation, and persistence (Hollenberg and van der Lubbe, 2011, p.12). The verbal score consists of tests regarding reading comprehension, vocabulary, writing, and spelling. The math score consists of tests on numbers and operations, proportions (incl. fractions and percentages), geometry, computations involving time and money, and associations.[4]

The main source of data on math and verbal skills is the national exam.[5] We received math and verbal scores from the national exam for 72% of pupils in our sample. These scores ranged from 0 to 100, with 50 corresponding to the nation wide average. Other schools provided us with the grades from the standardized student assessment system (developed by the same institution and with the same rationale as the national exam) called the "Leerlingvolgsysteem" (LVS; 18% of the sample). The LVS is less fine grained, distinguishing five different skill levels (A to E), with pupils in the lowest two skills brackets (D and E) on average differing by 15 percentage points. This effectively makes

[4] Although math and spatial reasoning are separate constructs of cognitive ability, they are both related to (and predictive of) educational outcomes and career choices in the STEM-fields, whereas verbal skills are mostly predictive of humanities and social sciences (Wai et al., 2009). Moreover, the math score in our sample is actually a composite math-spatial measure, as it also includes a geometry component, which is clearly linked to spatial ability.

[5] There are two types of national exams in the Netherlands. 61% of our sample completed the so-called "CITO"-test and 11% the so-called "Drempelonderzoek" (DO). These tests are comparable in nature, thus allowing to convert the scores to put them on a common basis.

a 15 (percentage) point difference the smallest unit of distinction for the test scores and we therefore also used this cutoff when we classified people on the basis of the national exam.

Individually balanced (IB) individuals were thus those children with an at most 15 point difference between math and verbal scores on the national exam, or receiving the same grade (A to E) in the LVS for both skills. IU individuals were those with mathematical and verbal test scores which differed by more than 15 percentage points of each other or receiving different grades; IU-math have a comparative advantage in math, while IU-verbal individuals have a comparative advantage in verbal skills. Note again that this definition implies that the group of IU individuals also includes some pupils that are not very good in either of the two subjects. However, these individuals still have a sizable comparative advantage in one subject.[6] 10% of the schools (and pupils) did not provide any test scores (on time) and were removed from the sample.

[INSERT FIGURE 2 AROUND HERE]

The left hand side of Figure 2 shows the distribution of individual types that are classified based on the national exam. Individuals close to the diagonal are the IB individuals, while those in the upper left (lower right) triangle are the IU-math (IU-verbal) pupils. (The graph in the middle panel of Figure 2 draws a similar scatter plot for the corresponding teams and is discussed below.) Our overall sample consisted of 428 IB, 219 IU-math and 187 IU-verbal subjects.

4.2 Team composition and randomization

In the second step of our two step procedure, we sorted the individual types *within a class* into four different team types. IB individuals were (in principle) always assigned to an IB team; see again Figure 1 for an illustration. IB teams therefore have balanced skills both at the individual and at the team level. In allocating IB individuals over different IB teams, we tried to keep these IB teams as comparable as possible in terms

[6] Indeed, the results from educational psychology confirm that students differentiate their educational and career development according to their relatively strongest skill (Lubinski and Benbow, 2006).

of average overall skills and gender composition.[7] We then randomly assigned the individually unbalanced subjects either into teams in which either mathematical skills or verbal skills dominate, or teams that consist of a combination of the two IU-types. Thus IU-M teams are composed of only IU-math individuals and IU-V teams consist only of IU-verbal individuals: these teams have a relatively high math or verbal score, both at the individual as well as at the team level. The IU teams in contrast, combining IU-math and IU-verbal subjects, have comparable scores in math and verbal skills at the team level, but not at the individual level. Again we kept a keen eye towards keeping average overall skills, and if possible gender composition, comparable. For each class the exact possibilities depended on the number of children (on average 24) and the distribution of individual types. Most teams consisted of five or six pupils. In a number of classes the limited degrees of freedom forced us to form a leftover team consisting of combinations of individual types that could not be classified as any of the team types of interest (these 17 leftover teams were omitted from the analysis). The graph in the middle panel of Figure 2 gives the distribution of team types corresponding to the pupils (and thus classes) classified based on the national exam.

Our design ensured that teams within the same class on average have (almost) the same overall skill level at the team level. The right hand panel of Figure 2 illustrates this for one particular class that contains six teams. The teams within this class differ in their combination of average math and verbal scores, but are close to each other in terms of average overall scores; the latter are reflected by the team dots on the 45 degree line (note that these average overall scores are close to each other and to the average overall skill level of 54.7 in this particular class). Obviously, average overall skill levels may differ substantially *between* classes; teams from different classes may thus be incomparable in this regard. In our empirical analysis we therefore only compare teams within the same class that compete for the same prize.

The teachers were not informed about the details and the purpose of the team composition. They were just informed of the (exogenous) team compositions with the explicit

[7] A pilot study we conducted in the year prior to this experiment revealed that girls have a comparative advantage in the production process (of friendship bracelets, key cords, etc.).

remark that teams should not be changed, unless they had strong objections against the team assignment (and not without our prior consent). Based on the teachers' objections, 12 children moved teams prior to the start of the program for (inter)personal reasons. As a result, three teams remained of the same type and three teams were no longer usable and were removed from the analysis.

Data on team performance were obtained from the teachers, who filled out a standardized spreadsheet that we already made them familiar with during the 'train the trainer' sessions. An example of the teacher spreadsheet is shown in Figure A1 in the Appendix. The teachers were instructed to provide information both during and at the end of the program, registering all transactions made by the teams, such as number of shares sold, share price, revenues etc. The excel sheet was constructed in such a way that the teachers could only fill out information in certain fields (marked in yellow). The excel sheet was programmed in such a way that with this information it automatically calculated the total amounts of money spent and received for each team per day, as well as the final score of the teams in Value of own shares (marked in blue) at the end of the program.[8] This score was then used to determine the winning team and the ranking of teams within the class. We demanded and closely monitored the use of this spreadsheet. A researcher visited each school at the end of the education program to confirm that no changes had been made to the initial team composition and to answer any questions concerning the excel sheet. The response rate for the completed excel files was 85.8%. The classes for which we did not receive the completed spreadsheet are excluded from our sample.

The final sample of teams for which all information was complete comprised 112 teams, covering 641 pupils spread over 29 classes: 60 IB teams, 25 IU teams, 14 IU-M teams and 13 IU-V teams. Besides collecting the spreadsheet and test scores from the schools, we obtained information directly from pupils by means of two questionnaires, one pre- and one post-treatment (the overall response rate for both questionnaires was 87%). The pre-treatment questionnaire asked a wide variety of questions about individual background characteristics, such as *Age*, whether *Female*, intended future high school track and it

[8]All other information (white fields) was password protected and could only be changed by the teacher after consultation with the research team.

measured entrepreneurial non-cognitive skills by means of a self-assessment. The latter consisted of nine different skills that have been found to be important for entrepreneurial activity, like creativity, risk taking propensity, persistence, etc. (see Huber et al. (2014) for more details). For the purpose of this study we calculate the unweighted average of these nine skills at the individual level. Intended future high school track ranged from 1 (pre-vocational secondary education) to 5 (pre-university education) and can be seen as a crude proxy for general ability (beyond math and verbal skills) and motivation / ambition. The post- treatment questionnaire was used to collect some information at the team level, such as the role that each student fulfilled within the team.

Panel A of Table 2 provides the descriptives of the team characteristics by team type, while panel B does so for class size and number of teams per class (as not all team types are represented in exact equal proportion in every single class, these numbers vary slightly over the different team types). The final column reports the p-values from ANOVA comparisons of the means of the various characteristics across the four team types. Given that our randomization took place within classes, the ANOVA comparisons for the team characteristics include class-id as a covariate to control for this. Hence the reported $p-$value is exactly equivalent to the one of testing for the joint significance of the four team type dummies in a (class) fixed effects regression of the characteristic at hand.[9] For the two class characteristics Numbers of Teams per Class and Class Size in panel B, class-id is not controlled for in the ANOVA (as by definition there is no within class variation in these characteristics over team types). There the ANOVA is thus based on between class variation, with the insignificance indicating that the four team types do not seem to be differently distributed over classes with different sizes and different numbers of teams. However, given the small sample size we interpret this finding with caution.

As intended given our experimental design, the teams differ in terms of average mathematical and verbal skills: IU-M teams have a (significantly) higher math score and IU-V teams a (significantly) higher verbal score compared to the other teams. Teams also differ

[9]Note that, as there is a one to one correspondence between classes and teachers, controlling for class-id is the same as teacher fixed effects.

in their average overall skills. A large part of this is due to the unavoidable variation between different classes; some schools/classes simply attract pupils with higher skills than others. Assessing the between class variation in total ability shows that there are significant differences in ability between different classes participating in our experiment (F-test = 13.19; p-value = 0.00). As explained above, this between class variation does not hamper our empirical analysis, as teams from different classes are not compared to each other. Unintentionally, however, there is also some within class variation in average overall skills; controlling for class-id in our ANOVA shows there are still (marginally) significant differences in total ability between team types within the same class (F-test = 2.27; p-value = 0.09 as reported in Table 2). The main culprit appears to be one class (with highly skilled pupils overall) that contains an IB team with exceptionally high verbal and math skills. Incidentally, this is the same class in which the team composition of three teams was changed at the teacher's request. Dropping this class reduces the sample to 108 teams for which no *within class differences* in average overall skills between team types are observed (F-test = 1.89; p-value = 0.14); this is reported in Table 2 with a separate line for Total Ability for $N = 108$.[10] In our analyses we consider both the full sample of 112 teams and the sub-sample of 108 teams (and also control for average overall skills in our class fixed effects regressions).

Moreover, in line with nationwide averages, girls in our sample score higher on verbal skills and boys score higher on mathematical skills. Therefore, despite our efforts to create balanced teams in terms of gender composition, the average share of females is significantly higher in IU-V teams and significantly lower in IU-M teams. A similar observation holds for intended future high school track. We will also control for these differences in our estimations. No differences are found in terms of average age and fraction of pupils that have at least one non-Dutch parent (as a proxy for ethnic differences).

[INSERT TABLE 2 AROUND HERE]

[10] For $N = 108$ a significant between class variation in total ability remains ($p = 0.00$ in the ANOVA). The descriptives and mean comparisons for the other characteristics are hardly affected when the sample is reduced from 112 to 108 teams and therefore not separately reported.

4.3 Outcome variables

The BizWorld program sets a clear and measurable objective for the participating teams, i.e., acquiring the highest ranking within one's class. The ranking is based on the financial measure *Value of Own Shares.* This was calculated by the teacher (in the excel sheet) as the total company value multiplied by the fraction of shares still owned by the team, i.e., not 'sold' to the 'venture capitalist' (i.e., the teacher). We use *Value of Own Shares* as one of the outcome measures. Note though that the value of own shares per se is insufficiently informative about how a team performed in the competition; for this one needs to know how it compares to the other teams. Our second performance measure is therefore *Normalized Team Rank.* On average there were 4.89 teams per class, with a minimum of two and a maximum of eight teams per class. Hence, team ranking can vary between one and eight, 1 being the best and 8 being the worst. Because it is easier to win in a class with fewer teams than in a class with more teams, we divided the rank of each team by the number of teams in its class, yielding the *Normalized Team Rank* variable.

Money won in the Tournament is the third, alternative, measure of team performance. The ratio of the prize money that can be won by the members of the winning team and the members of the team that comes in second place is 3:2. There was no money to be won by any of the other teams in the class. Hence, we assigned the value three (3) to all the wining teams, the value two (2) to the second best teams and zero (0) to all the remaining teams. An advantage of this outcome measure is that it takes the incentives to become first or second in the tournament into account. However, it does not take into account the non-financial benefits of ranking highly, nor does it control for the size of the competition, i.e., the number of teams in the class.

The descriptive statistics of the three outcome measures are shown in the lower panel of Table 2. Comparison of means across team types using ANOVA (controlling for class id) yields significant differences for all three of them. Table 3 reports the correlation matrix for all variables in Table 2, for ease of exposition leaving the class characteristics out.[11] Even though the assignment to the teams (i.e., the treatment) was random, there

[11]We only find significant correlations between class size on the one hand and % female, age and

are a number of variables that could potentially influence the effect of the team types on the outcome variables. For example, gender, age and ethnic background are demographic background characteristics that have been hypothesized and found to be important in the entrepreneurial team literature (Chowdhury, 2005; Jin et al., 2017). These are therefore included as covariates when analyzing the performance differences between the team types in more detail in the next section. Moreover, as discussed above, total ability and intended future high school track are also included as covariates to make sure that our findings not are driven by any remaining (marginal) differences between teams in average overall skills or other dimensions like ambition or motivation. Finally, because non-cognitive skills have also been shown to be important for entrepreneurial performance (Bosma et al., 2004; Aldén et al., 2017), we add average levels of non-cognitive skills at the team level as a control variable to our analysis.

[INSERT TABLE 3 AROUND HERE]

5 Results

5.1 Test of main hypotheses

To compare the mean performance outcomes across the various team types while controlling for covariates, we report in Table 4 the results from (class) fixed effect regressions for the three dependent variables of interest. To simplify the interpretation of the estimated regression coefficients, we have multiplied the value of Normalized Team Rank by -1. In all columns it thus holds that a positive coefficient implies a positive impact on performance.

The independent variables of main interest are the three team dummies IU-M team, IU-V team and IU-team. The coefficients belonging to these dummies reflect the mean difference in performance of these teams relative to the IB teams, which we use as a benchmark. The inclusion of class fixed effects ensures that estimated coefficients are based on within class variation only; differences in the tournament composition (as well

(logically) number of teams per class on the other hand.

as ability levels) between classes are thus fully accounted for. Apart from that we include a number of team characteristics from Table 2 as covariates, i.e., average age within the team, percentage of females in the team, the fraction of pupils of which one parent is not Dutch, team size, average intended future high school track, average non-cognitive skill level and average total ability score. The inclusion of the latter is another robustness check that remaining (minor) differences in average total ability across teams that compete against each other within the same class are not driving our results.[12] The observations are clustered at the class level to obtain robust standard errors (in parentheses) to account for the possibly correlated performance of the teams within one class.

[INSERT TABLE 4 AROUND HERE]

For each of the three performance measures we report the estimates for the full sample (columns (1), (3) and (5)), as well as for the sub-sample in which we exclude the class with an extremely high skilled IB-team and the changed teams at the teacher' request (columns (2), (4) and (6)). For each of the six specifications the three final rows in the table report the $p-$values of Wald tests that the estimated coefficients of two team dummies are the same. For instance, the row IU=IU-M tests for mean differences between the IU-M and IU teams, et cetera.

The results in all columns indicate a consistent pattern: IB teams perform significantly better than IU-V teams in terms of all three outcome variables, with no significant difference between IB and IU-M teams. Furthermore, the performance of the IB teams is significantly better than the performance of the IU teams. The latter finding provides strong support for Hypothesis 2. Thus, combining different types of individually unbalanced subjects within a team does not seem to substitute for composing teams of subjects that all have individually balanced skills.

The evidence for Hypothesis 1 is far less clear cut. That IB teams outperform IU-V teams is in line with it, but the equal performance of IU-M teams is not. Moreover, with decreasing returns to skills one would also expect that IU teams, which have balanced skills at the team level, outperform both IU-M and IU-V teams. This is not what we

[12] Controlling for average mathematical and verbal scores separately leads to the same main findings.

find, however. If anything, IU-M perform better than the IU teams, although differences are insignificant. Overall we thus find (at best) limited support for Hypothesis 1.[13] We thus take the observed difference between IB and IU teams as the main finding of our experiment. In the next subsection we provide an exploratory analysis of the potential mechanism underlying this result.

5.2 Exploration of the underlying mechanism

As discussed in Subsection 2.2, one potential reason why IB teams perform better than IU teams might be that team composition facilitates the effective assignment of tasks. From the loosely described responsibilities, as well as the CEO and the CFO being the only two team members involved in the presentation of the business plan and the selling of shares (cf. Section 3), especially these two roles seem key to team success. To explore potential cognitive skills related differences in task allocation, Table 5 first displays how the different skills sets are distributed among the members of winning and non-winning teams.[14]

[INSERT TABLES 5 AND 6 AROUND HERE]

As before, math and verbal scores are based on the grades (A=5 to E=1) from the student tracking system (LVS) and the converted nationwide exam scores (CITO).[15] From the table it can be observed that, on average, the winning teams assigned the person with the highest *and* most balanced test scores to the role of CFO, while CEOs come second in terms of test scores (but not in terms of balancedness). For the non-winning teams this also appears to be the case, but here differences with the other roles are (much) less

[13]Strictly speaking testing Hypothesis 1 only entails comparing IB and IU teams together, with IU-M and IU-V teams together. This can be done by running regressions similar to those in Table 4, with the only difference being that the three team dummies are replaced by a single dummy, which equals 1 iff a team is either of the IB or the IU type. For all six regressions displayed in Table 4 we then find an insignificant coefficient (the lowest p-value equals 0.155).

[14]The detailed information about the role assignment is based self-reported data collected in a post-treatment survey and only available for a sub-sample of the pupils. Together with the non-random assignment of roles, this makes the analysis presented here merely exploratory, leading to tentative conclusions at best.

[15]Due to the ex-post conversion from the more fine grained nationwide exam scores (1-100) to the grades (A-E) of the student tracking system, on some occasions students who had a 15-point or more difference in their test scores ended up in the same grade category after the conversion (e.g., a student with a test score of 51 in math and 68 in verbal was classified as IU-verbal type, yet in the ex-post conversion both test scores are classified as a 'B').

pronounced. Table 6 provides a similar overview of the average test scores by position in the team, but now does so for the four exogenously composed team types. Again, for all team types the CFO has the highest math and verb scores on average (while the CEO typically comes in second). However, in terms of balanced skills CFOs in IB teams also clearly rank highest among their fellow team members, while in IU teams this is not the case.[16] For example, within IU teams both the Director of Production as well as the Marketing Director have similar levels of skill balance on average. These descriptive statistics thus provide some first indication that having a CFO with both sufficient mathematical and sufficient verbal ability, i.e., a true master of all, might be beneficial for team performance.

Next we explore this explanation in a regression framework. To that purpose we first determined for each individual pupil whether he or she is a 'Master of All', which is defined as having an 'A' in both math and verbal test scores. Based on this classification of individuals, we created three dummies at the team level: 'Master of All in team', 'Master of All as CEO' and 'Master of All as CFO'. These respectively indicate whether there is a (i.e. at least one) master of all in the team, whether the CEO is a master of all, and whether the CFO is a master of all. (Clearly, the dummy 'Master of All in team' necessarily equals one whenever either 'Master of All as CEO' or 'Master of All as CFO' does so.) As to be expected, 57 out of the overall 60 IB teams have such a master of all within their ranks, while this only holds true for 11 out of the 25 IU teams.[17] Since we only observe the role assignment for a sub-sample of our teams (cf. Tables 5 and 6), the number of observations for the other two dummies (master of all as CEO / CFO) are substantially smaller. Among the 39 IB teams for which we do have observations, there are 25 with a master of all as CFO and 16 with a master of all as CEO (and 11 IB teams have both). For the observed 15 IU teams, only one has a master of all as CFO and none have a master of all as CEO. Likewise, also among the IU-M and IU-V teams there are

[16]Due to the limited number of observations for IU-V teams and IU-M teams, the descriptive statistics for especially these two team types should be viewed with caution. Together with the fact that the focus of this exploratory analysis is on IB and IU teams, this leads us to combine the IU-M and IU-V teams to one team type (IU-MV teams) in our analyses below.

[17]For the IU-M teams and the IU-V teams these numbers are 4 (out of 14) and 2 (out of 13), respectively.

only a few with a master of all as CFO, while masters of all as CEO are absent.[18]

[INSERT TABLES 7 AND 8 AROUND HERE]

Table 7 presents the results of adding - in various combinations - the master of all dummies to our main regression specification of Table 4. The results from these analyses clearly illustrate the significant positive relationship between having a master of all as CFO and team performance; while the 'Master of All in team' and 'Master of All as CEO' dummies are typically insignificant and occasionally significantly negative, the 'Master of All as CFO' dummy is always significantly positive.[19] To determine the effect of having a 'Master of All as CFO' compared to having no master of all at all in the team, we test for the significance of the linear combination of the estimated coefficients of the 'Master of All in team' and the 'Master of All as CFO' dummies and find that these are significant in all specifications (see the bottom rows in Table 7, where we also show the p-values for the linear combination involving 'Master of All as CEO').

Given that the CFO's task is highly oriented towards 'numbers', however, one could reasonably wonder whether strong verbal abilities are truely needed in this role. To make sure that it is indeed the balancedness of (high) skills and not just the high math ability of CFOs that matters for team performance, we conducted a similar analysis, but now also based on individuals being a 'Math Master' or not. A math master has an 'A' test score in math, but unlike a master of all, not necessarily so in his/her verbal test score. (A master of all is thus necessarily also a math master, but the reverse does not hold.) The results reported in Table 8 provide some suggestive evidence that balancedness of the CFO's skills is key; the "Math Master as CFO" dummy is never significantly positive. This suggests that having a Math Master in the role of CFO does not contribute much beyond the added value of having a Math Master in the team. Given that IB teams are much more likely than IU teams to have such a master of all within the team, the likelihood of having this person in the right role is also much higher for this team type.

[18]Specifically, for the IU-M teams and the IU-V teams the number of masters of all as CFO are 2 (out of 10) and 1 (out of 4), respectively.

[19]Performing the same analyses without control variables or excluding the class with the extremely high skilled IB-team leads to qualitatively similar results and does not change the conclusions drawn here.

Hence, the exploratory results presented here provide some evidence that this advantage in task allocation could be the underlying driver of the superior performance of IB teams compared to IU teams.

6 Concluding Discussion

6.1 Summary of the findings

Entrepreneurs are no longer lone wolves; nowadays, a significant share of successful ventures is started up by teams (Kamm et al., 1990). This has sparked growing interest in how team composition can be used as a strategic tool to enhance venture performance (West, 2007; Klotz et al., 2014; Jin et al., 2017). In particular, a burgeoning stream of research has begun to explore the impact on venture performance of socio-demographic team diversity in terms of age, education, wealth and gender (Hoogendoorn et al., 2013; Steffens et al., 2012; Colombo and Grilli, 2005; Chowdhury, 2005; Ruef et al., 2003). The present study has analyzed another, arguably more fundamental, dimension of team diversity: namely, cognitive skill composition. We have explored how verbal and mathematical skills, two key cognitive skills that are determinants of work performance in general, and which entrepreneurs also rely on, can best be combined in entrepreneurial teams.

Prior research has demonstrated that cognitive skills enhance the performance of *individual* entrepreneurs (Van der Sluis et al., 2008; Hartog et al., 2010). It has also shown that being a 'Jack of all Trades', i.e., possessing balanced skills (Lazear, 2005), is beneficial too. Our study is novel in exploring how these cognitive skills can best be combined and balanced *at the team level* in an entrepreneurial setting. We tested two hypotheses about optimal skill composition in teams in a randomized field experiment using data from the BizWorld entrepreneurship education program among 11-12 years old children. We found some support for the hypothesis that balanced cognitive skills promote venture performance in a team setting, but only if the skill balance comes from the combination of individuals that each have within-person skill balance. Combining team members with different (specialized) skills in mixed teams cannot replicate the superior performance of teams comprising members who individually possess balanced cognitive skills.

We believe our study contributes to the entrepreneurship literature by conducting a field experiment with exogenous variation in team composition, enabling sound causal inference. Even though team composition in real entrepreneurial teams is not random, exogenous variation is essential to be able to identify causal links between certain team characteristics and team outcomes (Jin et al., 2017, p.761). Our particular field experiment has some other attractive features, besides a number of limitations that will be discussed in the next subsection. First, it provides uniform, precise and objective measures of two important cognitive skills. Furthermore, the experiment is relatively large and run using young (school-age) subjects. The latter characteristic enables a relatively clean measurement of the impact of skills, because these subjects lack (heterogeneous) labor market experience and have not self-selected into particular study tracks. They all have to rely on their cognitive skills based on the same choices and investments. Moreover, teams formed within the context of entrepreneurship education programs are often assigned by the teacher who is aware of the skill sets of the individuals in his/her class. Hence, understanding how team composition affect the performance of the teams that are formed as an integral part of such programs, will provide valuable information to entrepreneurship education program designers and educators that might enhance entrepreneurship education effectiveness. The usefulness of our study for entrepreneurship education designers is further underlined by the fact that, unlike previous research on entrepreneurship education which has largely focused on intentions to become an entrepreneur, we focus on (objectively measured) actual team performance as outcome of main interest.

6.2 Limitations

Like any other empirical study, ours comes with a number of first order limitations common to carefully controlled, but thereby also tightly delineated, field experiments. Field experiments typically trade-off (higher) internal validity against (lower) external validity. Our experiment delivers interesting and novel insights into the causal relationship between team composition and team performance, in an interesting but specialized setting using a specific sample, specific incentive structure and a specific task structure that features many characteristics of entrepreneurial team work. However, this comes at the cost of limited generalizability.

First, in terms of sample, our experimental subjects are 11-12 years old pupils setting up toy firms in a specific entrepreneurship education program. Besides the potential advantages of using this age group as discussed above, this choice also implies some limitations. For one thing, children may not yet have completed the development of their mental abilities (Baltes et al., 1999; Burt, 1954). More importantly, the impact of team composition might be substantially moderated by skills regarding team organization and coordination. Such skills are likely to be developed at schools and universities and also later on at work. The performance penalty from mixing specialists that we find, relative to a team of members with balanced skills, might thus be less pronounced in a sample of an older population. Based on the existing literature, there is as yet no basis for presuming that the *direction* of the treatment effects we find for pupils should be any different for older subjects. Nevertheless, while we believe that our approach provides an interesting first step towards testing our hypotheses and carries some important insights for teams performing entrepreneurial tasks in an educational setting, it may not fully translate to the real-world entrepreneurial context.

Second, in line with actual entrepreneurial settings where one should do better than the competition in order to be successful and survive, we employed an incentive structure in which only the best two teams within a class received a (fixed and given) monetary reward. This may have given teams incentives to take risks, an aspect that we consider representative of entrepreneurial ventures as well. Obviously, however, given that we only studied this particular competitive incentive scheme, we cannot exclude that different results would be obtained under different incentive structures. For instance, team behavior and thus outcomes might potentially be different under a more gradual incentive scheme in which improvements in rank below the second best place, or improvements in share performance that do not immediately lead to a higher rank, are also rewarded. A related concern might be that the stakes involved are potentially too low, thus providing pupils with little incentive to do their best. As noted in Section 3, the financial prizes pupils could win correspond to about one or two weeks of 'salary' (i.e., pocket money at that age), which does not seem negligible. Moreover, other field experiments have found that (even unexpected) sheer recognition increases subsequent performance, and especially so when it is exclusively provided to the best performers (Bradler et al., 2016). Together with the awards and gifts ('recognition') provided by the program and the entrepreneur, the financial stakes thus seems to provide considerable incentives to exert effort in order to win the competition. Apart from that, there seem little a priori reason to believe that (too) low stakes drive our finding that IB teams outperform IU teams.

A third limitation concerns the specific task structure used in the BizWorld program. Although team members are assigned a specific position in the team - and with that are given some loosely described responsibilities - the program is clearly geared towards stimulating team work. Tasks are thus highly interdependent, with all team members (to various degrees) being involved in all company tasks. For instance, the manufacturing of the friendship bracelets that the teams produce is done by all team members together. Although this setup might well be representative of entrepreneurial teams, especially early on in the entrepreneurial venture, it might potentially also favor IB teams. In particular, the interdependence of tasks might require more collaboration and communication in order to build consensus and engender coordination among team members (Straus, 1999; Wageman, 1995; McGrath, 1984). These features, may not be pertinent to other entrepreneurial settings where team members can specialize in particular tasks and hence (mixed) specialists can be better employed (Wageman, 1995).

Other task related limitations of the BizWorld program concern their scope, duration and the nature of the product made. In terms of scope the tasks do not involve opportunity recognition, for instance, which is clearly important in entrepreneurial practice. The short duration of the program also hinders the external validity of our experiment, because interactions and communication between team members may in practice alter over time due to a better understanding of each other, irritations that arise, or due to routines that evolve (cf. Webber and Donahue, 2001; Bell et al., 2011). Finally, the task structure is also specific in the exact combination of skills required. The fact that the friendship bracelets are technologically simple may have favored having balanced skills, as no truly expert knowledge is needed to produce these. This may well be different in actual practice, especially for startups in the high tech industry. Nevertheless, also in our setting some cognitive skills appear more valuable than others. The good performance of the IU-M teams suggests that overall (i.e. at the level of the entire team) math skills are more valuable than verbal skills in the BizWorld competition.

The present study focused only on mathematical and verbal skills: in doing so, it may have ignored other relevant (cognitive) skills. For various reasons we feel that this is not a major limitation, though. Studies from psychology (see Cattell, 1963 and Cattell and Horn, 1978) have demonstrated that the general factor of intelligence is comprised of three distinct cognitive abilities, which can also be classified as verbal, mathematical, and spatial ability (Wai et al., 2009, p.821). As noted in Subsection 4.1 the math score in our sample is in fact a composite measure of mathematical and spatial ability. Although this precludes us from using a genuine measure of balancedness in verbal, mathematical and spatial ability, the measure that we employ should be closely related as it encompasses all these abilities. Moreover, although an exact, one-to-one comparison is hard to make, our skills measures are very similar to the constructs developed and used to predict high school grades and future fields of education and occupation in several psychological studies (Gustafsson and Balke, 1993; Wai et al., 2009).[20] Importantly, these studies examined the additional explanatory value of mathematical and verbal skills *beyond* general intelligence for academic achievement (Deary et al., 2007). Overall, findings from these studies suggest that domain-specific achievement factors and specific aptitude factors measured at the age of 11–13 are important predictors of educational performance and occupational choices

[20]In the economics of education literature math and verbal skills are also often used as key outcome measures to assess the impact of educational interventions (see e.g., Angrist et al., 2010, 2012). Furthermore, when explaining his Jack-of-all-trades theory, Lazear (2005) himself uses math and verbal skills as an example of the two key raw skills in his model (see p. 676-677).

in closely related domains later in life. All in all we are therefore confident that the math and verbal test scores we used as well as the balance between them provide us with objectively and precisely measured proxies for the cognitive skill dimensions that are predictive of future cognitive skills and educational outcomes, which in turn influence entrepreneurial performance (Gustafsson and Balke, 1993; Shea et al., 2001; Wai et al., 2009).

References

Aghion, P. and Howitt, P. (2006). Joseph schumpeter lecture appropriate growth policy: A unifying framework. *Journal of the European Economic Association*, 4(2-3):269–314.

Aldén, L., Hammarstedt, M., and Neuman, E. (2017). All about balance? a test of the jack-of-all-trades theory using military enlistment data. *Labour Economics*, 49:1–13.

Aldrich, H. E. and Fiol, C. M. (1994). Fools rush in? the institutional context of industry creation. *Academy of Management Review*, 19(4):645–670.

Allison, T. H., Davis, B. C., Short, J. C., and Webb, J. W. (2015). Crowdfunding in a prosocial microlending environment: Examining the role of intrinsic versus extrinsic cues. *Entrepreneurship Theory and Practice*, 39(1):53–73.

Almus, M. and Nerlinger, E. A. (1999). Growth of new technology-based firms: which factors matter? *Small Business Economics*, 13(2):141–154.

Angrist, J. D., Dynarski, S. M., Kane, T. J., Pathak, P. A., and Walters, C. R. (2010). Inputs and impacts in charter schools: Kipp lynn. *American Economic Review*, 100(2):239–43.

Angrist, J. D., Dynarski, S. M., Kane, T. J., Pathak, P. A., and Walters, C. R. (2012). Who benefits from kipp? *Journal of policy Analysis and Management*, 31(4):837–860.

Åstebro, T. and Serrano, C. J. (2015). Business partners: Complementary assets, financing, and invention commercialization. *Journal of Economics & Management Strategy*, 24(2):228–252.

Åstebro, T. and Thompson, P. (2011). Entrepreneurs, jacks of all trades or hobos? *Research Policy*, 40(5):637–649.

Åstebro, T. and Yong, K. (2016). Invention quality and entrepreneurial earnings: the role of prior employment variety. *Entrepreneurship Theory and Practice*, 40(2):381–400.

Baltes, P. B., Staudinger, U. M., and Lindenberger, U. (1999). Lifespan psychology: Theory and application to intellectual functioning. *Annual review of psychology*, 50(1):471–507.

Baron, R. A. and Henry, R. A. (2010). How entrepreneurs acquire the capacity to excel: Insights from research on expert performance. *Strategic Entrepreneurship Journal*, 4(1):49–65.

Bell, S. T., Villado, A. J., Lukasik, M. A., Belau, L., and Briggs, A. L. (2011). Getting specific about demographic diversity variable and team performance relationships: A meta-analysis. *Journal of management*, 37(3):709–743.

Bosma, N., Van Praag, M., Thurik, R., and De Wit, G. (2004). The value of human and social capital investments for the business performance of startups. *Small Business Economics*, 23(3):227–236.

Brannon, D. L., Wiklund, J., and Haynie, J. M. (2013). The varying effects of family relationships in entrepreneurial teams. *Entrepreneurship Theory and Practice*, 37(1):107–132.

Burt, C. (1954). The differentiation of intellectual ability. *British Journal of Educational Psychology*, 24(2):76–90.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54(1):1.

Cattell, R. B. and Horn, J. L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurement*, 15(3):139–164.

Chen, L.-W. and Thompson, P. (2016). Skill balance and entrepreneurship evidence from online career histories. *Entrepreneurship Theory and Practice*, 40(2):289–305.

Chowdhury, S. (2005). Demographic diversity for building an effective entrepreneurial team: is it important? *Journal of Business Venturing*, 20(6):727–746.

Colombo, M. and Grilli, L. (2005). Founders human capital and the growth of new technology-based firms: A competence-based view. *Research Policy*, 34(6):795–816.

Custódio, C., Ferreira, M. A., and Matos, P. (2013). Generalists versus specialists: Lifetime work experience and chief executive officer pay. *Journal of Financial Economics*, 108(2):471–492.

Deary, I. J., Strand, S., Smith, P., and Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35(1):13–21.

Eesley, C. E., Hsu, D. H., and Roberts, E. B. (2014). The contingent effects of top management teams on venture performance: Aligning founding team composition with innovation strategy and commercialization environment. *Strategic Management Journal*, 35(12):1798–1817.

Elert, N., Andersson, F. W., and Wennberg, K. (2015). The impact of entrepreneurship education in high school on long-term entrepreneurial performance. *Journal of Economic Behavior & Organization*, 111:209–223.

Ferreira, D. and Sah, R. K. (2012). Who gets to the top? generalists versus specialists in managerial organizations. *The RAND Journal of Economics*, 43(4):577–601.

Foo, M.-D. (2011). Teams developing business ideas: how member characteristics and conflict affect member-rated team effectiveness. *Small Business Economics*, 36(1):33–46.

Gao, H., Yu, T., and Cannella Jr, A. A. (2016). The use of public language in strategy: A multidisciplinary review and research agenda. *Journal of Management*, 42(1):21–54.

Gustafsson, J.-E. and Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, 28(4):407–434.

Hartog, J., Van Praag, M., and Van Der Sluis, J. (2010). If you are so smart, why aren't you an entrepreneur? returns to cognitive and social ability: Entrepreneurs versus employees. *Journal of Economics & Management Strategy*, 19(4):947–989.

Hollenberg, J. and van der Lubbe, M. (2011). Toetsen op school primair onderwijs. http://www.cito.nl/onderzoek

Hoogendoorn, S., Oosterbeek, H., and Van Praag, C. (2013). The impact of gender diversity on the performance of business teams: Evidence from a field experiment. *Management Science*, 59(7):1514–1528.

Hoogendoorn, S., Parker, S. C., and Van Praag, M. (2017). Smart or diverse start-up teams? evidence from a field experiment. *Organization Science*, 28(6):1010–1028.

Horwitz, S. K. and Horwitz, I. B. (2007). The effects of team diversity on team outcomes: A meta-analytic review of team demography. *Journal of Management*, 33(6):987–1015.

Howell, J. M. and Higgins, C. A. (1990). Champions of technological innovation. *Administrative Science Quarterly*, pages 317–341.

Huber, L. R., Sloof, R., and Van Praag, M. (2014). The effect of early entrepreneurship education: Evidence from a field experiment. *European Economic Review*, 72:76–97.

Jin, L., Madison, K., Kraiczy, N. D., Kellermanns, F. W., Crook, T. R., and Xi, J. (2017). Entrepreneurial team composition characteristics and new venture performance: A meta-analysis. *Entrepreneurship Theory and Practice*, 41(5):743–771.

Kacperczyk, A. and Younkin, P. (2017). The paradox of breadth: The tension between experience and legitimacy in the transition to entrepreneurship. *Administrative Science Quarterly*, 62(4):731–764.

Kaiser, U. and Müller, B. (2015). Skill heterogeneity in startups and its development over time. *Small Business Economics*, 45(4):787–804.

Kamm, J., Shuman, J., Seeger, J., and Nurick, A. (1990). Entrepreneurial teams in new venture creation: A research agenda. *Entrepreneurship: Theory and Practice*, 14(4):7–17.

Kim, P. H., Aldrich, H. E., et al. (2005). Social capital and entrepreneurship. *Foundations and Trends® in Entrepreneurship*, 1(2):55–104.

Kim, P. H. and Longest, K. C. (2014). You can't leave your work behind: Employment experience and founding collaborations. *Journal of Business Venturing*, 29(6):785–806.

Klotz, A., Hmieleski, K., Bradley, B., and Busenitz, L. (2014). New venture teams a review of the literature and roadmap for future research. *Journal of Management*, 40(1):226–255.

Kollmann, T., Stöckmann, C., Meves, Y., and Kensbock, J. M. (2017). When members of entrepreneurial teams differ: linking diversity in individual-level entrepreneurial orientation to team performance. *Small Business Economics*, 48(4):843–859.

Kremer, M. (1993). The o-ring theory of economic development. *Quarterly Journal of Economics*, pages 551–575.

Kroll, M., Walters, B. A., and Le, S. A. (2007). The impact of board composition and top management team ownership structure on post-ipo performance in young entrepreneurial firms. *Academy of Management Journal*, 50(5):1198–1216.

Lazear, E. (2005). Entrepreneurship. *Journal of Labor Economics*, 23(4):649–680.

Lechmann, D. S. and Schnabel, C. (2014). Are the self-employed really jacks-of-all-trades? testing the assumptions and implications of lazears theory of entrepreneurship with german data. *Small Business Economics*, 42(1):59–76.

Lounsbury, M. and Glynn, M. A. (2001). Cultural entrepreneurship: Stories, legitimacy, and the acquisition of resources. *Strategic Management Journal*, 22(6-7):545–564.

Lubinski, D. and Benbow, C. P. (2006). Study of mathematically precocious youth after 35 years: Uncovering antecedents for the development of math-science expertise. *Perspectives on Psychological Science*, 1(4):316–345.

Martens, M. L., Jennings, J. E., and Jennings, P. D. (2007). Do the stories they tell get them the money they need? the role of entrepreneurial narratives in resource acquisition. *Academy of Management Journal*, 50(5):1107–1132.

Martin, B. C., McNally, J. J., and Kay, M. J. (2013). Examining the formation of human capital in entrepreneurship: A meta-analysis of entrepreneurship education outcomes. *Journal of Business Venturing*, 28(2):211–224.

Marvel, M. R., Davis, J. L., and Sproul, C. R. (2016). Human capital and entrepreneurship research: A critical review and future directions. *Entrepreneurship Theory and Practice*, 40(3):599–626.

McGrath, J. E. (1984). *Groups: Interaction and performance.* Englewood Cliffs, NJ: Prentice-Hall.

Mueller, P. (2006). Exploring the knowledge filter: How entrepreneurship and university-industry relationships drive economic growth. *Research Policy*, 35(10):1499–1508.

Oosterbeek, H., van Praag, M., and Ijsselstein, A. (2010). The impact of entrepreneurship education on entrepreneurship skills and motivation. *European Economic Review*, 54(3):442–454.

Parhankangas, A. and Ehrlich, M. (2014). How entrepreneurs seduce business angels: An impression management approach. *Journal of Business Venturing*, 29(4):543–564.

Parker, S. C. (2009). Can cognitive biases explain venture team homophily? *Strategic Entrepreneurship Journal*, 3(1):67–83.

Patel, P. C. and Ganzach, Y. (2019). Returns to balance in cognitive skills for the self-employed: evidence from 18 countries. *Small Business Economics*, 52(1):89–109.

Premand, P., Brodmann, S., Almeida, R., Grun, R., and Barouni, M. (2016). Entrepreneurship education and entry into self-employment among university graduates. *World Development*, 77:311–327.

Ruef, M., Aldrich, H., and Carter, N. (2003). The structure of founding teams: Homophily, strong ties, and isolation among u.s. entrepreneurs. *American Sociological Review*, 68(2):195–222.

Shea, D. L., Lubinski, D., and Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, 93(3):604.

Shephard, R. W. (2012). *Cost and production functions*, volume 194. Springer Science & Business Media.

Steffens, P., Terjesen, S., and Davidsson, P. (2012). Birds of a feather get lost together: new venture team composition and performance. *Small Business Economics*, 39(3):727–743.

Straus, S. G. (1999). Testing a typology of tasks: An empirical validation of mcgrath's (1984) group task circumplex. *Small Group Research*, 30(2):166–187.

Strohmeyer, R., Tonoyan, V., and Jennings, J. E. (2017). Jacks-(and jills)-of-all-trades: On whether, how and why gender influences firm innovativeness. *Journal of Business Venturing*, 32(5):498–518.

Ucbasaran, D., Lockett, A., Wright, M., and Westhead, P. (2003). Entrepreneurial founder teams: Factors associated with member entry and exit. *Entrepreneurship Theory and Practice*, 28(2):107–128.

Unger, J., Rauch, A., Frese, M., and Rosenbusch, N. (2011). Human capital and entrepreneurial success: A meta-analytical review. *Journal of Business Venturing*, 26(3):341–358.

Van der Sluis, J., Van Praag, M., and Vijverberg, W. (2008). Education and entrepreneurship selection and performance: A review of the empirical literature. *Journal of economic surveys*, 22(5):795–841.

Van Praag, C. and Cramer, J. (2001). The roots of entrepreneurship and labour demand: Individual ability and low risk aversion. *Economica*, 68(269):45–62.

Vanaelst, I., Clarysse, B., Wright, M., Lockett, A., Moray, N., and S'Jegers, R. (2006). Entrepreneurial team development in academic spinouts: An examination of team heterogeneity. *Entrepreneurship Theory and Practice*, 30(2):249–271.

Von Graevenitz, G., Harhoff, D., and Weber, R. (2010). The effects of entrepreneurship education. *Journal of Economic Behavior & Organization*, 76(1):90–112.

Wageman, R. (1995). Interdependence and group effectiveness. *Administrative science quarterly*, pages 145–180.

Wagner, J. (2006). Are nascent entrepreneurs jacks-of-all-trades? a test of lazear's theory of entrepreneurship with german data. *Applied Economics*, 38(20):2415–2419.

Wai, J., Lubinski, D., and Benbow, C. P. (2009). Spatial ability for stem domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101(4):817.

Webber, S. S. and Donahue, L. M. (2001). Impact of highly and less job-related diversity on work group cohesion and performance: A meta-analysis. *Journal of management*, 27(2):141–162.

West, G. (2007). Collective cognition: When entrepreneurial teams, not individuals, make decisions. *Entrepreneurship Theory and Practice*, 31(1):77–102.

Wright, M., Hmieleski, K. M., Siegel, D. S., and Ensley, M. D. (2007). The role of human capital in technological entrepreneurship. *Entrepreneurship Theory and Practice*, 31(6):791–806.

Zheng, Y., Devaughn, M. L., and Zellmer-Bruhn, M. (2016). Shared and shared alike? f ounders' prior shared experience and performance of newly founded banks. *Strategic Management Journal*, 37(12):2503–2520.

Figure 1: Team composition: from individuals to teams

Combination of skills (per person) Math	Verbal			Individual type		in Team type
H	H	= IB_H	}	Individually balanced (IB)	→	IB team
M	M	= IB_M				
L	L	= IB_L				
H	M	= $Math_{HM}$	}	Individually unbalanced (IU)-math	→	IU-M team
H	L	= $Math_{HL}$			→	IU team
M	L	= $Math_{ML}$				
M	H	= $Verb_{MH}$	}	Individually unbalanced (IU)-verbal	→	IU team
L	H	= $Verb_{LH}$			→	IU-V team
L	M	= $Verb_{LM}$				

*Note: For illustrative purposes we divided each skill into three possible levels: high (H), medium (M), and low (L). In reality this is a continuum (0-100) of possible values.

Figure 2: Distribution of Individual and Team types

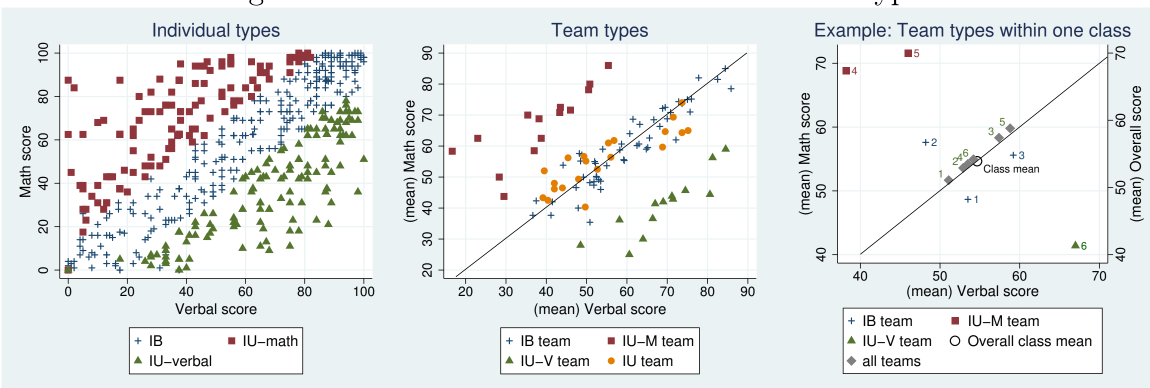

Table 1: The BizWorld program

Day 1	Introduction and theory on entrepreneurship
	Apply for position in team
Day 2	Register company and receive 10 shares
	Present business plan to "venture capitalist" to raise start-up capital
	Company stock prices displayed in class
Day 3	Design and manufacture products (friendship bracelets)
	Calculate production costs (incl. rent, material, salaries, etc.)
	Determine product prices
Day 4	Design marketing campaign (poster and "commercial")
	Sell products to pupils in lower grade
Day 5	Complete profit- and loss statement and balance sheet
	Winning team announced and rewarded

Table 2: Descriptives by team type

	Total sample		IB teams		IU teams		IU-M teams		IU-V teams		ANOVA	
	mean	std. dev.	mean	std. dev.	mean	std. dev.	mean	std. dev.	mean	std. dev.	F-test	p-value
A: Team Background characteristics												
Math score	3.72	0.57	3.83	0.56	3.62	0.40	4.12	0.46	3.03	0.40	35.00	0.00
Verbal score	3.77	0.63	3.91	0.55	3.64	0.54	2.97	0.57	4.24	0.40	30.96	0.00
Total ability score (N=112)	7.49	1.01	7.74	1.07	7.26	0.85	7.09	0.98	7.27	0.74	2.27	0.09
Total ability score (N=108)	*7.43*	*0.97*	*7.65*	*1.03*	*7.20*	*0.83*	*7.09*	*0.98*	*7.27*	*0.74*	*1.89.*	*0.14*
Non cognitive skills	4.58	0.33	4.62	0.33	4.58	0.36	4.51	0.33	4.65	0.30	0.10	0.96
% Female	0.49	0.15	0.51	0.14	0.45	0.11	0.37	0.16	0.58	0.19	6.43	0.00
Age	11.72	0.36	11.68	0.36	11.71	0.31	11.89	0.36	11.72	0.42	1.33	0.27
High school track	3.11	0.68	3.28	0.66	2.88	0.72	3.00	0.57	2.94	0.62	6.79	0.00
Team size	5.72	0.59	5.75	0.63	5.88	0.44	5.43	0.51	5.62	0.65	0.36	0.78
Nat. parents: One non-dutch	0.14	0.16	0.15	0.16	0.12	0.15	0.09	0.12	0.20	0.22	1.42	0.24
B: Class characteristics												
Number of teams per class	4.89	1.30	4.92	1.25	5.04	1.31	4.36	1.08	5.08	1.66	1.00	0.40
Class size	25.17	4.53	25.37	4.43	26.08	4.05	23.86	6.11	23.92	3.84	1.10	0.35
C: Outcome variables												
Value of own shares	109.26	101.40	122.37	112.52	91.28	81.49	130.57	104.41	60.38	55.85	3.01	0.03
Normalized Team Rank	0.57	0.27	0.52	0.27	0.65	0.27	0.50	0.20	0.73	0.22	2.91	0.04
Money won in Tournament	1.16	1.30	1.35	1.36	0.84	1.18	1.71	1.20	0.31	0.75	2.97	0.04
Number of teams	112		60		25		14		13			

*Note: Average math and verbal score are based on the grades (A=5 to E=1) from the student tracking system (LVS) and the converted nationwide exam scores (CITO). Total ability score is based on the sum of the average math and verbal scores. High school track refers to the intended future high school track and ranges from 1 (pre-vocational secondary education) to 5 (pre-university education). Normalized Team Rank is the rank per team divided by the number of teams in its class. Money won in Tournament is the ratio of the prize that can be won by the winning team, the runner up and all the other teams (3:2:0).

Table 3: Correlation matrix

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(1) IB team	1.00															
(2) IU-M team	-0.41***	1.00														
(3) IU-V team	-0.39***	-0.14	1.00													
(4) IU team	-0.58***	-0.20**	-0.19**	1.00												
(5) Math score	0.19**	0.26***	-0.44***	-0.10	1.00											
(6) Verbal score	0.24**	-0.48***	0.27***	-0.11	0.39	1.00										
(7) Total ability score	0.26**	-0.15	-0.08	-0.13	0.82***	0.85***	1.00									
(8) Non cognitive skills	0.06	-0.11	0.06	-0.03	0.14	0.27***	0.25***	1.00								
(9) % Female	0.18*	-0.31***	0.22**	-0.14	-0.06	0.26**	0.13	-0.02	1.00							
(10) Age	-0.11	0.18*	-0.00	-0.01	-0.02	-0.18*	-0.13	0.09	-0.00	1.00						
(11) High school track	0.26	-0.06	-0.09	-0.19*	0.50***	0.56***	0.64***	0.29***	0.05	-0.19**	1.00					
(12) Team size	0.05	-0.19**	-0.07	0.14	-0.07	0.03	-0.02	0.01	-0.01	-0.15	0.08	1.00				
(13) Nat. parents: One non-dutch	0.05	-0.11	0.14	-0.08	-0.15	0.04	-0.06	0.16*	0.08	0.18*	-0.03	0.03	1.00			
(14) Normalized Team Rank	0.20**	0.11	-0.22**	-0.16*	0.10	-0.14	-0.03	0.01	-0.11	0.20	-0.02	0.20**	0.05	1.00		
(15) Value of own shares	0.14	0.08	-0.18*	-0.10	0.11	-0.03	0.04	0.10	-0.21**	0.11	0.11	0.16*	-0.10	0.61***	1.00	
(16) Money won	0.16*	0.16*	-0.24**	-0.13	0.09	-0.19**	-0.07	-0.02	-0.12	0.17*	-0.06	0.15	-0.02	0.85***	0.61***	1.00

*Note: N=112; */**/*** indicates a significance at the 10%/5%/1%-level.

Table 4: The effect of team type on team performance

	Value of own shares		Normalized Team Rank		Money won in Tournament	
	(1)	(2)	(3)	(4)	(5)	(6)
Team dummies *(omitted category: IB team)*						
IU-M team	-17.07	-13.67	-0.04	-0.03	0.01	0.06
	(30.03)	(30.40)	(0.11)	(0.11)	(0.58)	(0.58)
IU-V team	-62.16***	-61.41***	-0.23**	-0.23**	-1.30***	-1.24***
	(20.65)	(21.61)	(0.10)	(0.10)	(0.42)	(0.43)
IU team	-57.73**	-52.20**	-0.15**	-0.13*	-0.59**	-0.52*
	(23.61)	(24.08)	(0.07)	(0.07)	(0.29)	(0.29)
Team characteristics						
Age	95.35***	92.64***	0.37***	0.36***	1.31***	1.23**
	(28.79)	(28.58)	(0.10)	(0.10)	(0.47)	(0.47)
Female	-72.79	-63.84	-0.08	-0.04	0.27	0.35
	(69.00)	(71.35)	(0.24)	(0.24)	(1.17)	(1.19)
Nationality	-139.77**	-135.01**	-0.30	-0.29	-0.69	-0.60
	(61.86)	(61.81)	(0.20)	(0.20)	(1.09)	(1.11)
Team size	63.30***	66.58***	0.18***	0.19***	0.62**	0.63**
	(17.36)	(19.66)	(0.04)	(0.05)	(0.25)	(0.29)
High school track	-23.69	-23.93	-0.09	-0.09	-0.60	-0.49
	(24.57)	(26.31)	(0.09)	(0.10)	(0.47)	(0.51)
Non-Cognitive skills	26.02	24.36	0.09	0.09	0.39	0.38
	(28.65)	(29.21)	(0.10)	(0.10)	(0.54)	(0.55)
Total ability score	23.91	23.02	0.10*	0.09	0.62*	0.61*
	(21.57)	(22.87)	(0.06)	(0.06)	(0.31)	(0.31)
Number of teams	112	108	112	108	112	108
R^2 (within)	0.29	0.27	0.28	0.27	0.22	0.20
p-values of comparing team dummies						
IU-M = IU-V	0.21	0.20	0.22	0.20	0.07	0.07
IU-M = IU	0.18	0.21	0.35	0.39	0.32	0.34
IU-V = IU	0.88	0.77	0.50	0.41	0.12	0.13

* Note: */**/*** indicates a significance at the 10%/5%/1%-level. Observations are clustered at the class level. Robust standard errors in parentheses. High school track refers to the intended future high school track and ranges from 1 (pre-vocational secondary education) to 5 (pre-university education). Nationality is measured by the fraction of pupils of which one parent is not Dutch. Total ability score is based on the sum of the average math and verbal scores (based on the grades, A=5 to E=1) from the student tracking system (LVS) and the converted nationwide exam scores (CITO). To simplify the interpretation of the estimated coefficients for Normalized Team Rank, we multiply these values by -1. IU = IU-M, IU = IU-V, IU-M = IU-V, IU = IU-M = IU-V refer to p-values from Wald-tests comparing the estimated coefficients between the specified team types.

Table 5: Winners vs. non-winners: average test scores by position in team

Test score	Winner			Non-winner		
	math	verb	|diff. M-V|	math	verb	|diff. M-V|
CEO	4.13	4.07	0.73	3.91	3.77	0.77
	(1.13)	(1.03)	(0.70)	(1.20)	(1.23)	(0.87)
CFO	4.53	4.40	0.40	4.34	4.09	0.62
	(0.74)	(1.18)	(0.63)	(1.00)	(1.10)	(0.71)
Design	3.14	3.43	0.57	3.33	3.47	0.94
	(1.23)	(1.02)	(0.76)	(1.34)	(1.08)	(0.77)
Production	3.21	3.07	0.71	3.58	3.57	0.59
	(1.42)	(1.49)	(0.83)	(1.18)	(1.38)	(0.65)
Marketing	4.09	3.45	0.63	3.32	3.56	0.80
	(0.94)	(0.93)	(0.81)	(1.33)	(1.16)	(0.67)
Sales	3.57	3.07	0.79	3.29	3.19	0.78
	(1.22)	(1.27)	(0.80)	(1.40)	(1.36)	(0.96)
Number of teams		15			53	

*Note: Standard deviations in parentheses. The detailed information about the role assignment is based on self-reported data collected in a post-treatment survey and only available for a sub-sample of the students. Math and verbal score are based on the grades (A=5 to E=1) from the student tracking system (LVS) and the converted nationwide exam scores (CITO). The absolute difference between math and verbal scores is calculated at the individual level.

Table 6: Average test scores by position in the team by team type

	IB teams			IU teams			IU-M teams			IU-V teams		
	math	verb	|diff. M-V|	math	verb	|diff. M-V|	math	verb	|diff. M-V|	math	verb	|diff. M-V|
CEO	4.13	4.23	0.26	3.80	3.60	1.40	4.10	2.60	1.50	2.50	4.00	1.50
	(1.17)	(1.01)	(0.44)	(1.15)	(1.24)	(0.74)	(0.99)	(0.97)	(0.71)	(1.00)	(1.15)	(1.00)
CFO	4.41	4.44	0.23	4.27	3.80	1.13	4.50	3.40	1.10	4.25	4.75	0.50
	(1.12)	(1.07)	(0.54)	(0.70)	(1.01)	(0.52)	(0.71)	(1.17)	(0.74)	(0.50)	(0.50)	(0.58)
Design	3.31	3.51	0.37	3.14	3.64	1.57	4.00	3.00	1.20	1.75	3.5	1.75
	(1.21)	(1.17)	(0.55)	(1.35)	(0.74)	(0.51)	(1.33)	(0.94)	(0.42)	(0.96)	(1.29)	(0.96)
Production	3.36	3.42	0.28	3.62	3.69	1.15	4.00	2.71	1.29	3.50	4.25	0.75
	(1.47)	(1.50)	(0.45)	(0.77)	(1.38)	(0.80)	(0.82)	(1.11)	(0.49)	(0.58)	(0.50)	(0.50)
Marketing	3.63	3.64	0.53	3.46	3.84	1.15	3.33	2.33	1.00	2.50	3.50	1.00
	(1.38)	(1.14)	(0.62)	(1.39)	(1.07)	(0.80)	(0.52)	(0.52)	(0.63)	(0.58)	(0.58)	(0.00)
Sales	3.33	3.39	0.30	3.23	2.85	1.77	3.86	2.71	1.14	3.00	3.00	1.00
	(1.45)	(1.41)	(0.53)	(1.42)	(1.34)	(0.83)	(0.69)	(0.95)	(1.07)	(1.41)	(0.00)	(0.00)
Number of teams		39			15			10			4	

*Note: Standard deviations in parentheses. The detailed information about the role assignment is based on self-reported data collected in a post-treatment survey and only available for a sub-sample of the students. Math and verbal score are based on the grades (A=5 to E=1) from the student tracking system (LVS) and the converted nationwide exam scores (CITO). The absolute difference between math and verbal scores is calculated at the individual level.

Table 7: The effect of 'Masters of All' on team performance

	Value of own shares				Normalized Team Rank				Money won in Tournament			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Team dummies *(omitted category: IB team)*												
IU-MV team	-31.84	43.97	28.64	28.28	-0.17*	0.06	-0.01	-0.01	-0.60	0.13	-0.21	-0.22
	(28.89)	(44.11)	(38.82)	(41.12)	(0.10)	(0.16)	(0.11)	(0.12)	(0.50)	(0.87)	(0.61)	(0.64)
IU team	-50.46*	-4.11	0.97	-2.79	-0.16**	-0.02	0.01	-0.02	-0.51	0.25	0.37	0.28
	(28.07)	(41.85)	(45.19)	(44.40)	(0.08)	(0.12)	(0.12)	(0.12)	(0.34)	(0.62)	(0.63)	(0.65)
Master of all dummies												
Master of All in team	15.06	65.67	-4.33	8.37	-0.03	0.17	-0.16	-0.07	0.17	0.81	-0.79	-0.49
	(26.04)	(40.54)	(31.38)	(37.20)	(0.09)	(0.13)	(0.09)	(0.12)	(0.44)	(0.76)	(0.55)	(0.66)
Master of All as CEO		-36.41		-28.17		-0.24**		-0.21*		-0.87		-0.68
		(32.02)		(32.15)		(0.10)		(0.11)		(0.54)		(0.51)
Master of All as CFO			79.71***	76.49**			0.35***	0.32***			1.81***	1.73***
			(26.63)	(27.50)			(0.08)	(0.09)			(0.45)	(0.51)
Team controls included	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Number of teams	112	68	68	68	112	68	68	68	112	68	68	68
R^2 (within)	0.28	0.27	0.34	0.35	0.26	0.31	0.38	0.44	0.18	0.19	0.32	0.35
p-values of linear combinations of Master of All dummies												
MoA_team + MoA_CEO = 0		0.53		0.59		0.62		0.04		0.95		0.11
MoA_team + MoA_CFO = 0			0.06	0.05			0.05	0.02			0.03	0.04

* Note: */**/*** indicates a significance at the 10%/5%/1%-level. Observations are clustered at the class level. Robust standard errors in parentheses. 'Master of all in team' is a dummy variable indicating if there is at least one master of all in the team. Masters of all are defined as having an 'A' in both math and verbal test scores. 'Masters of All as CFO (CEO)' is a dummy variable which takes on the value 1 if the CFO (CEO) in a team is a 'Masters of All'. Team controls included are the same as in our main analysis shown in Table 4. To simplify the interpretation of the estimated coefficients for Normalized Team Rank, we multiply these values by -1. MoA_team + MoA_CEO = 0 and MoA_team + MoA_CFO = 0 refer to the p-values from Wald-tests for the linear combination of the estimated coefficients of 'Master of All in team' + 'Master of All as CEO (CFO)'.

Table 8: The effect of 'Math Masters' on team performance

	Value of own shares		Normalized Team Rank		Money won in Tournament	
	(1)	(2)	(3)	(4)	(5)	(6)
Team dummies *(omitted category: IB team)*						
IU-MV team	-33.74	12.83	-0.08	0.07	-0.19	0.25
	(24.88)	(29.94)	(0.09)	(0.12)	(0.47)	(0.58)
IU team	-58.12**	-25.20	-0.16**	-0.05	-0.62*	0.03
	(23.93)	(38.14)	(0.07)	(0.10)	(0.31)	(0.44)
Math Master dummies						
Math Master in team	17.90	-1.62	0.14	0.26*	1.14***	1.61**
	(33.08)	(55.05)	(0.09)	(0.13)	(0.36)	(0.68)
Math Master as CFO		31.00		-0.02		-0.02
		(21.85)		(0.10)		(0.49)
Team controls included	YES	YES	YES	YES	YES	YES
Number of teams	112	68	112	68	112	68
R^2 (within)	0.28	0.25	0.28	0.29	0.23	0.24
p-values of linear combinations of Math Master dummies						
MM_team + MM_CFO = 0		0.56		0.09		0.02

* Note: */**/*** indicates a significance at the 10%/5%/1%-level. Observations are clustered at the class level. Robust standard errors in parentheses. 'Math Master in team' is a dummy variable indicating if there is at least one math master in the team. Math Masters are defined as having an 'A' in their math test score but not necessarily in their verbal test score. 'Math Master as CFO' is a dummy variable which takes on the value 1 if the CFO in a team is a 'Math Master'. Team controls included are the same as in our main analysis shown in Table 4. To simplify the interpretation of the estimated coefficients for Normalized Team Rank, we multiply these values by -1. MM_team + MM_CFO = 0 refer to the p-values from Wald-tests for the linear combination of the estimated coefficients of 'Math Master in team' + 'Math Master as CFO'.

Figure A1: Example of teacher spread sheet

BizWorld register Class number: 269	School name: Example school	Name trainer: John Smith	Name Teacher: Magret Jones	BizWorld dates: June 2, 4, 9 and 11

day 1			Money received				Money spent							Cash final	Sustainability:			Team mission
Company/ Team name:	Team number	cash in register	# shares sold	Price per share	Other revenues	Total	Design kit	Expenses yarn	Other expenses	Patent (5 or 0)	Company registration	Salaries & rent	Total	Start cash + receiv. - spent	total # yarn bought	of which # sustainable yarn	Sustainability trade mark (yes/no)	CSR in mission (yes/no)
Friends & Co.	269 1	1	2	14		29	4	3		5	1	7	20	9	3	1	no	yes
V and B	269 2	1	2	11		23	4	5		5	1	6	21	2	5	0	no	no
Made by Uzz!	269 3	1	2	12	2	27	4	4		5	1	7	21	6	4	2	yes	yes

day 2			Money received					Money spent						Cash final	Sustainability:		
Company/ Team name:	Team number	cash in register	# shares sold	Price per share	Loan	Other revenues	Total	Expenses yarn	Other expenses	Packaging mat. (2 BE)	Salaries & rent	Interest on loan	Total	Start cash + receiv. - spent	Packaging material bought (yes/no)	total # yarn bought	of which # sustainable yarn
Friends & Co.	269 1	9		8	10	2	12	5	1	0	7	1	14	7	no	5	2
V and B	269 2	2	1	8	10		18	4		2	7	1	14	6	yes	8	0
Made by Uzz!	269 3	6		8	10		10	4	4	0	7	1	16	0	no	8	4

day 3			Money received					Money spent				Cash final	Sustainability:	Results production:		buyers	
Company/ Team name:	Team number	cash in register	# shares sold	Price per share	Loan	Revenues Big Sale	Total	Marketing Pack.	Salaries & rent	Interest on loan	Total	Start cash + receiv. - spent	CSR bonus points (max. 50 points per team)	# bracelets made	# bracelets sold	total # BE exchanged	(1 BE = € 0,20)
Friends & Co.	269 1	7	1	18		97	115	10	7	1	18	104	30	31	29	36	€ 2,40
V and B	269 2	6	1	16		55	71	10	7	1	18	59	0	19	19		
Made by Uzz!	269 3	0	1	20		82	102	10	7	1	18	84	10	21	21		

day 4			Money received			Cash Final	Profit and loss statement					Final score of the team			
Company/ Team name:	Team number	cash in register	Salaries & rent	Interest on loan	Loan redeemed	Cash before taxes	Revenues	Total expenses	Profit before taxes	Taxes 30%	Net Profit	**Total company value**	# own shares	**Value of own shares**	number of team members
Friends & Co.	269 1	104	7	1	10	86	97	60	37	11	26	260	7	182	6
V and B	269 2	59	7	1	10	41	55	61	-6	0	0	0	6	0	6
Made by Uzz!	269 3	84	7	1	10	66	82	63	19	5	14	140	7	98	6