## Entrepreneurship, teams and sustainability: A series of field experiments

Rosendahl Huber, L.

**Publication date**
2015
**Document Version**
Final published version

This dissertation reports the results from three field experiments that were conducted within the setting of one of the leading, internationally renowned entrepreneurship education programs for primary schools called BizWorld. The first field experiment evaluates the program's effectiveness in terms of the development of entrepreneurship knowledge and a set of non-cognitive skills relevant for entrepreneurial activity. The results indicate that the program has a robust positive effect on non-cognitive entrepreneurial skills. The aim of the second experiment is to test how skill composition affects team performance and whether (a lack of) individual balanced skills can be substituted by combining the skills of various specialists within one team. The results show that balanced skills are beneficial to team performance, and that it is hard to effectively combine different specialists within one team. The third experiment investigates how to induce sustainable behavior in a productive setting. The results indicate that, in this setting, (financial) incentives are required to effectively motivate sustainable behavior.

Laura Rosendahl Huber (1979) obtained her bachelor's and master's degree in Business Economics at the University of Amsterdam. In 2010 she started her PhD at the Amsterdam Center for Entrepreneurship at the University of Amsterdam. Her main research interests are entrepreneurship, education and organizational economics. After her PhD she will join the Max Planck Intitute for Innovation and Competition in Munich as a Senior Research Fellow.

Entrepreneurship, teams and sustainability: a series of field experiments
Laura Rosendahl Huber

610

Laura Rosendahl Huber

# Entrepreneurship, teams and sustainability: a series of field experiments

Universiteit van Amsterdam

# ENTREPRENEURSHIP, TEAMS AND SUSTAINABILITY: A SERIES OF FIELD EXPERIMENTS

ISBN 978 90 361 0432 6

Cover design: Crasborn Graphic Designers bno, Valkenburg a/d Geul

This book is no. 610 of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

# ENTREPRENEURSHIP, TEAMS AND SUSTAINABILITY: A SERIES OF FIELD EXPERIMENTS

**ACADEMISCH PROEFSCHRIFT**

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus

prof. dr. D.C. van den Boom

ten overstaan van een door het college voor promoties

ingestelde commissie,

in het openbaar te verdedigen in de Agnietenkapel

op donderdag 16 april 2015, te 10:00 uur

door

**Laura Rosendahl Huber**

geboren te Eindhoven

# PROMOTIECOMMISSIE

**Promotores:**

Prof. dr. C.M. van Praag, Copenhagen Business School en Universiteit van Amsterdam
Prof. dr. R. Sloof, Universiteit van Amsterdam

**Overige leden:**

Dr. A.M. Franco, University of Toronto
Prof. dr. D. Harhoff, Max Planck Institute for Innovation and Competition
Prof. dr. P.D. Koellinger, Universiteit van Amsterdam
Prof. dr. E.J.S. Plug, Universiteit van Amsterdam
Prof. dr. A.H.G. Rinnooy Kan, Universiteit van Amsterdam
Dr. J. van de Ven, Universiteit van Amsterdam

Faculteit Economie en Bedrijfskunde

# Acknowledgements

Wow. I can't believe that it is (almost) done. Four years of work wrapped up together in one book. I would like to thank all my friends and my family for all the help and support over the past few years. There are a few people that I would like to thank in particular. First and foremost, my two supervisors. Mirjam, your (silent) persistence and your endless enthusiasm about research brought me (back) to academia. Thank you for your caring support through all the ups and downs that are part of research and part of life. Randolph, you have the great gift of providing critical comments and asking critical questions in the most friendly and confidence building way. Together you are a perfect team. Always one step ahead, and even though you are very different in terms of character and research topics, you somehow manage to agree on almost everything (most of the time). I would like to thank you both for all your time, effort, inspiration and wisdom. I could not have wished for better supervisors.

I would also like to thank my paranymphs, Inez and Eszter. I do not think I would have made it through TI without you and I sincerely hope that our (working) paths will cross again sometime in the future. Inez, I will miss our weekly lunches. Eszter, thank you for the talks, walks and the fun trip to Bergen. And even though I am only allowed to have two paranymphs, I would also like to thank Sander and Adam as my unofficial paranymphs. Thank you both for the "inspiring" conference visits, dinners and exhausting bike rides. Silvia, thank you for the coffee breaks at times when I needed them most. Martin, Joeri, and Thomas it was a pleasure to have you as my colleagues.

I would also like to thank Pauline van Dulken for trusting me and for allowing me to carry out these field experiments within the BizWorld program. The support of all the people at the Amsterdam Center for Entrepreneurship and of Erik Boer in particular is also gratefully acknowledged.

Finally, I would like to thank Jip and Sophie. Thank you for your confidence and love. Thank you for your patience, especially during the last few months. You believed in my ability to complete this task long before I realized you were right. I am looking forward to our new adventure in Germany and to whatever comes after that.

*Laura Rosendahl Huber*

*Amsterdam, January 2015*

# Contents

# Chapter 1

# Introduction

## 1.1 Motivation

The aim of this dissertation is to answer three research questions: (1) Can entrepreneurship be taught (in school, when young)?, (2) What is the effect of balanced skills on team performance?, and (3) Can incentives induce sustainable behavior?

The first question has been the subject of discussion for many years (e.g., Lindquist et al., 2013; Colombier and Masclet, 2008). The sharp increase in the number of entrepreneurship education programs suggests that the general consensus is that entrepreneurship can indeed be taught. From a policy perspective this is an appealing thought. The idea that entrepreneurs are not necessarily born but can also be developed creates a window of opportunity for (educational) policies aimed at enhancing entrepreneurship. However, there is little research on the effectiveness of such educational programs. Chapter 3 of this dissertation evaluates the effectiveness of an early entrepreneurship education program. A theoretical motivation to look at *early* entrepreneurship education is provided by Cunha and Heckman's (2007) general model of the technology of skill formation. This model emphasizes the importance of early investments in both cognitive and non-cognitive skills. It strongly suggests that an investment in skills not only has a direct impact on the current stock of skills, but also produces spill-over effects in subsequent periods by boosting current skills and by making investments later in life more productive. Early investments in skills may thus be particularly effective in the long run. Obviously, the (potential) future spill-over benefits of early investments in skills only occur if the early investment has a direct impact on the stock of skills in the first place. Chapter 3 of this dissertation therefore evaluates the *direct* (short term) effect of early entrepreneurship education.

The second question is motivated by the observation that nowadays, teamwork is an omnipresent phenomenon within firms of all types and sizes (Hamilton et al., 2003). Large firms increasingly rely on the work and decisions made by (self-managed) teams

(Lazear and Shaw, 2007). Moreover, a substantial and growing share of businesses are started up and run by entrepreneurial teams instead of solo entrepreneurs (Klotz et al., 2014; Parker, 2009). Over the past decades the tasks within all types of organizations (from new ventures to established firms) have become increasingly complex due to new technologies and rapidly changing environments (Dahlin et al., 2005; Lazear and Shaw, 2007). The combination of an increase in task complexity and the observed increase in teamwork raises the question of effective team composition. This question, focusing on the *skill* composition of successful (entrepreneurial) teams, is empirically explored in Chapter 4 of this dissertation.

The third question, which will be addressed in Chapter 5, is motivated by the economics literature related to pro-social behavior. A set of recent field experiments shows that financial and non-financial incentives have a positive effect on social behavior (Ashraf et al., 2014; Miller et al., 2012; Olken et al., 2012). These experiments are all conducted in settings in which social behavior (related to health and educational outcomes) is the main aim of the program and thus the main outcome variable of interest. However, social behavior has also become a topic of interest for many (large) for-profit firms, and social goals are sometimes added to traditional financial goals (Kitzmueller and Shimshack, 2012). Within these companies a method has to be found to effectively balance these two (possibly conflicting) goals. A potential solution to induce sustainable behavior could be through the use of incentives. Various studies have looked at the association between incentives and corporate social performance, and the empirical evidence is rather mixed (see Walls et al. (2012) for an overview). The drawback of many empirical studies is that they do not account for a non-random, i.e., endogenous, matching between a CEO and a firm, and between a firm and its remuneration policy. Hence, it is virtually impossible to establish a causal link when studying the association between CEO incentives and CSR using only observational data. The experimental design described in Chapter 5 solves these methodological difficulties while studying the effect of incentives on social behavior in a setting in which financial performance also matters.

The link between these three questions is primarily given by the use of the same experimental setting and the same type of research method. The field experiments described in this dissertation were conducted within the setting of one of the leading, internationally renowned entrepreneurship education programs for primary schools (called BizWorld). The BizWorld program aims to teach children aged 11 or 12 the basics of business and entrepreneurship through an experiential learning program that takes five days (within a time span of 2 to 4 weeks). At the start of the program, the class is divided into teams of five or six children. Within each team, each team fulfills his/her specific role (e.g. CEO, CFO, Sales director, etc.) besides working together

as a team. During the lessons, all five with a practical orientation, the children set up a toy business in friendship bracelets and go through a firm's entire business cycle (from start-up to liquidation). More specifically, the teams have to: write and present a business plan in order to raise start-up capital, design and manufacture products (friendship bracelets), calculate production costs and determine product prices, sell the products during a sales market to the pupils in the grade below, and finally they have to complete a profit and loss statement. Individual team members have strong incentives to care about the business performance of their team. In the school year 2009-2010, when we conducted the field experiments on team composition and sustainability, team members of the winning team were awarded a gift voucher of €7,50 each, and the team members of the runner up were each awarded a gift voucher of €5,00. This is an addition to the certificate for the winning team that is provided by the BizWorld foundation.

## 1.2 Method and contribution

This dissertation reports the results from three field experiments. Field experiments are often used in medical trials, but up until recently their application in economics was scarce. In this dissertation this research method is applied to address the three research questions described above.

The aim of medical trials is typically to estimate a causal effect of a new drug on the treatment of a certain disease. Similarly, the field experiments described in this dissertation are also aimed at estimating the causal effect of a certain treatment variation on the outcome variables of interest. To be able to estimate a causal effect, a group of people (or subjects) is *randomly* divided into two or more subgroups, each with a certain treatment. As a result of the random treatment assignment, the subjects in the different groups are on average equal in terms of observed *and* unobserved characteristics. Hence, the only difference between these groups is in the treatment that they are exposed to. Thus, any differences that are observed in the average group outcomes can be attributed directly to the treatment.

In medical trials the treatment typically consists of the old medicine, the new medicine and a placebo, which serves as the control group. In this dissertation the treatments and the treatment groups vary in each chapter. Chapter 3 evaluates the effectiveness of the entrepreneurship education program. Thus, in this setting, the treatment is the entrepreneurship program itself and the control group (or the placebo) is the regular school curriculum. The outcome variables of interest in this chapter are the development of entrepreneurial skills and knowledge. The experiment described in Chapter 4 is about how team composition (in terms of skill balance) affects team per-

formance. In this chapter the treatment thus consists of different team compositions, i.e., children are randomly assigned to a certain team conditional on their individual skill set. The outcome variable of interest in this chapter is team performance. Chapter 5 tests the effect of incentives on sustainable behavior. The chapter reports the results of two treatments that are aimed at inducing this type of behavior. The outcome variable of interest is sustainable behavior and the teams in the regular course setting serve as the control group.

Each of the three topics described above has been studied before. A novelty and contribution of this dissertation is that the experimental design applied here enables the estimation of causal effects related to questions that are otherwise (often) troubled with methodological difficulties which hinder these causal inferences such as self-selection, omitted variable bias or reverse causality. This is relatively new in the domain that we study. For example, people who are more interested in entrepreneurship are more likely to enroll in an entrepreneurship education program. This selection into the program is likely to be correlated with unobserved characteristics such as an individual's motivation during the program and a person's initial entrepreneurial ability. Thus, simply comparing the entrepreneurial skills of people who chose to participate in an entrepreneurship education program with those who did not is likely to lead to an overestimation of the effect of the program (see Section 3.1). Moreover, new venture teams are seldom exogenously composed but are typically the result of the combination of the (unobserved) preferences of the individual team members. Thus, self-selection and omitted variable bias also hinder the estimation of causal effects of team composition on team performance when using observational data. Finally, as mentioned above, the choice to encourage sustainable behavior or to introduce incentives to induce sustainable behavior is often endogenous, i.e., driven by observed and unobserved individual and firm level characteristics, which makes it difficult to establish a causal link.

In addition to the estimation of causal effects, each chapter contributes to the economics literature in other ways. Chapter 3 is the first study to evaluate the effects of entrepreneurship education of children in primary school (ages 11 and 12). Previous studies of the impact of youth entrepreneurship education follow adolescents and mainly focused entrepreneurial intentions. Thus, another contribution of this chapter is the focus on the development of both knowledge and skills. The contribution of Chapter 4 is based on the fact that thus far there is little evidence on the effect of (balanced) skills at the team level on (team) performance. Another contribution of this chapter is that the experimental design creates the opportunity to explicitly study the intra-team substitutability of useful combinations of skills. Chapter 5 uses an experimental design to study sustainable behavior in a productive environment in which both sustainable behavior and financial performance matter. Furthermore, within this productive setting

the aim is to study sustainable behavior at the team level, instead of at the individual level. This is a realistic feature, since production decisions are often made by teams. Incentives might prove to be a more effective way to induce sustainable behavior in teams (rather than selection of motivated individuals that then need to be combined in a team in specific ways).

As always, and this dissertation forms no exception, each of the studies has some limitations as well. These will be discussed in each of the main chapters and in Chapter 6.

## 1.3 Outline

This dissertation is structured as follows. Chapter 2 describes the entrepreneurship education program that is used for the three field experiments and provides some information on the data collection.

Chapter 3 is based on the paper "The effect of early entrepreneurship education: Evidence from a field experiment", co-authored with Randolph Sloof and Mirjam Van Praag (European Economic Review 72, 2014, 76–97). The aim of this chapter is to analyze the effectiveness of *early* entrepreneurship education. To this end, a leading entrepreneurship education program that is taught worldwide in the final grade of primary school is evaluated. To assess the impact of the program we focus on pupils' development of entrepreneurship knowledge and a set of non-cognitive skills relevant for entrepreneurial activity. The results indicate that knowledge is unaffected by the program. However, we find that the program has a robust positive effect on seven (out of nine) non-cognitive entrepreneurial skills. Self-reported scores on (constructs of) Risk taking propensity, Creativity, Need for Achievement, Self-Efficacy, Pro-activity, Persistence and Analyzing all increase significantly more in the treatment group than in the control group. This is surprising since previous evaluations found zero or negative effects. Because these earlier studies all pertain to entrepreneurship education for adolescents, our result tentatively suggests that non-cognitive entrepreneurial skills are best developed at an early age.

Chapter 4 is based on the paper "Jacks-of-all-trades? The effect of balanced skills on team performance", co-authored with Randolph Sloof and Mirjam Van Praag. The aim of this chapter is to test how skill composition affects team performance and whether (a lack of) individual balanced skills can be substituted by combining the skills of various specialists within one team. This chapter is partly motivated by the findings presented by Rulke and Galaskiewicz (2000). They find that the teams of generalists outperform the specialist teams. We add to their study by (among other things) distinguishing between two specific *types* of specialists. Based on these types we

explicitly compose teams consisting of different combinations of specialists and compare their performance to the generalist teams. Based on pupils' precisely measured level of verbal and mathematical ability, we exogenously compose 179 teams separated into four different team types: JAT teams, math-specialist teams, verbal-specialist teams and mixed-specialist teams. Our results show that balanced skills are beneficial to team performance, and that it is hard to substitute individual balanced skills by combining different specialists within one team.

Chapter 5 is based on the paper "The effect of incentives on sustainable behavior: Evidence from a field experiment", also co-authored with Randolph Sloof and Mirjam Van Praag. This chapter investigates how to induce sustainable behavior in a productive setting and the link between sustainable behavior and financial performance. To this end, the schools participating in the entrepreneurship education program are randomly assigned to one of three treatments: the first is purely financially oriented, the second promotes sustainable behavior and the third also induces sustainability by incentives. Comparing the outcomes of the three groups in terms of sustainable behavior and financial performance the results imply that, in this setting, (financial) incentives are required to effectively motivate sustainable behavior. Moreover, the choice to behave more sustainable is not associated with financial performance.

Chapter 6 provides a brief summary of the previous chapters, some concluding remarks, and indicates some potential avenues for future research.

# Chapter 2

# Program and context

*This chapter describes the entrepreneurship education program that is used to conduct the field experiments which are discussed in Chapters 3 through 5 of this dissertation.*

## 2.1 Program

The entrepreneurship education program that is evaluated in Chapter 3 and that is used as the setting for the field experiments described in Chapter 4 and 5 is called BizWorld. It is one of the leading entrepreneurship education programs worldwide for primary schools.[1] The program originated in the United States in the late 1990's. Since its inception, over 500,000 children from more than 100 countries have participated in one of their education programs.[2]

The program consists of five teaching days which can be taught over the course of a 2 to 4 week period. It is a structured program and a day-by-day overview of the content is shown in Table 2.1. The lessons, all five with a practical orientation, lead the participating pupils through a firm's business cycle from start-up to liquidation. The first day starts with a theoretical introduction on entrepreneurship. At the start of the practical part on the first day, the teacher divides the class into teams of five or six children.[3] Each child then writes an application letter applying for his/her preferred role within their team. The positions to be fulfilled are: General Manager (CEO), Finance

[1] A similar international program is the 'Young Enterprise' program offered by the Junior Achievement Worldwide network.

[2] The BizWorld Foundation offers three different education programs for children from the third to the eighth grade (i.e. BizWorld, BizWiz and BizMovie). BizWorld is the largest with approximately two-third of all the children participating in this program (Source: www.bizworld.org/Bizworld-at-a-Glance).

[3] In the regular course of the program the team composition is determined by the teacher. However, during the school year of 2009-2010 we were allowed to compose the teams. The details will be discussed in Chapter 4.

**Table 2.1.** The BizWorld program

| | **The program (baseline)** | **CSR treatments** |
|---|---|---|
| ***Day 1*** | Introduction and theory on entrepreneurship | |
| | Apply for position in team | |
| ***Day 2*** | Register company and receive 10 shares | Explain reward structure |
| | Present business plan to "venture capitalist" to raise start-up capital | Explain and discuss CSR |
| | Company stock prices displayed in class | Encourage CSR in mission |
| ***Day 3*** | Introduce sustainability trademark and sustainable yarn | |
| | 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 | Depends on reward structure |

Director (CFO), Director of Product Design, Director of Manufacturing, Marketing Director, and Sales Director. The teacher matches the candidates to positions based on their knowledge of the child, the child's application letter and the job descriptions provided in the course guidelines. During the course of the program the team members fulfill their specific roles besides working (and learning) together as a team.

On the second day, the teams decide on a company name and officially register their company with the "chamber of commerce".[4] Next, all the teams write a business plan, which is presented to a "venture capitalist" in order to sell stocks and to raise start-up capital. The quality of the business plan, based on some specific characteristics described in the course guidelines, together with the presentation determine the share price the investor is willing to pay. Each team receives ten shares at the registration of their company. They can decide on the number of shares they want to sell to the investor. However, each team has to sell some shares, because they need some cash before they can start the design and production process. All transactions are made in 'BizEuros' instead of actual Euros.

The third day is devoted to design, procurement and production. The available raw materials for sale (see Figure 2.1) are most suitable for producing friendship bracelets, although bookmarks or key or phone cords are alternative possibilities. Production is prepared intensively because production time is limited (to one hour). After having calculated production costs, including salaries, raw materials and rent, the companies determine the sales price.

From the third day onwards, the teams have two alternative routes to raise more

[4]During the course of the program the children have to go to several official agencies, i.e., chamber of commerce, bank, venture capitalist, etc. These roles are played by the teacher.

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. It is explained to them that if they sell too many shares they lose ownership of the company. This is a difficult concept and the teacher and the entrepreneur try to convey as clearly as possible the importance of this factor in determining the winning team.

The fourth day is used for preparing the marketing campaign, which consists of a poster, the store presentation and a "commercial" (i.e., a two minute stage play). On this day, the products are also sold to the children in the grade below, usually at an organized fair. Before the sale starts, each team is given the opportunity to present their product by means of their "commercial" in front of the group of prospective buyers. After the sales market is over, 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 after having redeemed their loans and paid profit taxes. The financial report is a basic ingredient to assessing the performance of teams. 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 (Bradler et al., 2013). Firstly, the BizWorld foundation provides certificates for each member of the winning team. Furthermore, 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. Additionally, but only in the year 2009-2010, team members of the winning team were awarded a gift voucher of € 7,50 each, and the team members of the runner up were each awarded a gift voucher of € 5,00. The average amount of pocket money that children receive at that age is €4 per week or €15 per month.[5] So the incentives we provide are relatively strong. The teacher could choose between two types of gift certificates; one that could be used at all major chain stores in the Netherlands or one that could only be spent on books. The majority of the teachers ($\frac{2}{3}$) chose the general gift certificate, the other teachers chose the book voucher. Finally, based on our observations, BizWorld teams usually show a strong motivation to achieve good company results and win the competition in their class.

Another addition that was made to the program for the purpose of the experiment is the option to exchange money. During the sales market the buyers (i.e., the children from the grade below) are offered the option to exchange their BizEuro (BE) for real euros (exchange rate: 1BE= € 0,20). In the regular course set-up the amount of BEs

[5]This information was collected for the children in our sample through a question in the pre-test questionnaire.

each buyer receives is calculated according to the following formula: # of sellers*15/ # of buyers. Hence, without the possibility of exchanging BizEuros for real euros the average revenue per seller is expected to be equal across schools.[6] By allowing the children to exchange money we provide them with an alternative in case they do not want to spend all their BEs, e.g., because there are not enough products that they like or that they want to buy.[7] This extra option allows us to compare the financial performance of teams across schools in a meaningful way.

### 2.1.1 Sustainability

To explain sustainability to the children in our experiment, we focus on nature and the environment. This is introduced to the program in several ways. First, sustainable yarn is added to the course material for all the schools participating in the experiment (see Figure 2.1). This yarn is sold at the same price as the regular yarn. The "cost" attached to the use of this type of yarn is in the less attractive look and the loss of diversity in colors (i.e., the sustainable yarn was only available in one color, as opposed to eight different colors for the regular yarn). Secondly, before the teams start with the design of their product and the production of their prototype a sustainability trademark is introduced. The teams can apply for this trademark if they fulfill two distinct criteria: (1) at least 50% of the yarn used in the production of the bracelets has to be sustainable yarn, and (2) the plastic packaging material available in the course materials can not be used. Hence, the teams that obtain this trademark have to be more creative and have to find alternative ways to display their products. Teams are only allowed to use the trademark in their official communication, such as the commercial or the poster, if their prototype was approved by the teacher or the entrepreneur. Finally, the teacher or the entrepreneur could give points for sustainable behavior (e.g. if the team paid attention to the amount of waste produced).

## 2.2 Context

The course materials for the teacher, containing all the details about the education program, are provided by the program. The materials are handed out during a two hour train-the-trainer session a couple of weeks prior to the program. The guidelines for the program are very strict and described in detail in the instruction manual which is part of the course material. Additionally, instruction videos are available on the

[6] In practice there is quite some variation in the the revenue per seller across the different classes (mean = 15.64, SD = 5.50, min = 6.21 and max = 33.76).

[7] In 34% of the classes some BE were exchanged for real euros.

**Figure 2.1.** Course material and sustainable yarn

BizWorld website, to give the teacher a preview of the course content.[8]

The sample used in this dissertation includes schools in (the western part of) the Netherlands. The Dutch BizWorld program started in 2004 and over 30.000 children have since then participated.[9] Unlike in the United States, the Dutch program is taught by an entrepreneur (or someone from the business world) in cooperation with the teacher. The entrepreneur brings real life examples and experiences into the classroom. In The Netherlands, all classes in the last grade of all primary schools -whether private or public- are eligible for BizWorld. Schools usually get in touch with the program through BizWorld marketing campaigns (i.e., BizWorld sending letters to schools to invite them to participate) or through sponsoring entrepreneurs or companies (from the neighborhood for instance). In general the BizWorld Foundation matches schools and sponsoring entrepreneurs willing to participate. Furthermore, the Dutch program is externally funded (sponsored by companies and/or subsidized by the government) and is therefore free of charge for the schools. Thus, financial or network constraints do not hinder schools' participation in the program. Schools sign up for the program at the beginning of each school year (before January). Most schools have either one or two (parallel) classes in last grade. In general, the voluntary decision to participate is taken at the school level (for all classes in the last grade), although it is possible that one class in a school does participate, whereas the other does not. The minimum level of participation is an entire class, i.e., individual pupils or teams cannot participate.

## 2.3 Data collection

To be able to monitor each school closely, the experiments were conducted at schools in the western part of the Netherlands (close to where I live). The data collection for Chapter 3 was conducted during the school years 2009-2010 and 2010-2011. The data for Chapter 4 and Chapter 5 were collected in the school year 2009-2010. An overview of the treatment assignment in the school year of 2009-2010 is shown in Table

[8] See: www.bizworld.org/Teacher-Resources or http://www.jongondernemen.nl/bizworld

[9] In 2013 the BizWorld program in the Netherlands merged with the organization 'Jong Ondernemen'. The set-up and the content of the education program have remained the same.

A2.1 in the appendix. Because the treatment variations that are described in the different chapters of this dissertation are conducted simultaneously, several checks are preformed to confirm that there is no interference between the different treatments. These will be described in more detail in each of the following chapters.[10] To gather the required information on individual (background) characteristics, all pupils had to complete two extensive questionnaires, measuring not only entrepreneurship knowledge, skills and intentions but also a wide array of individual background characteristics. All the children had to fill out one questionnaire before and one after the program. Data on team performance are obtained via the teachers. They filled out a standardized spreadsheet during and at the end of the program to register all transactions made by the teams, such as number of shares sold, share price, revenues etc.

[10] A correlation matrix of the treatments is shown in Table A2.2 in the appendix.

# Appendix

Table A2.1: Treatment assignment matrix for school year 2009-2010

| | baseline | non-incent. CSR | incent. CSR | Total | | |
|---|---|---|---|---|---|---|
| JAT-team | 25 | 25 | 36 | 105 | 86 | treatment ent. educ. |
| | *5* | *6* | *8* | | *19* | *control ent. educ.* |
| Mixed-spec. team | 13 | 7 | 10 | 38 | 30 | treatment ent. educ. |
| | *3* | *2* | *3* | | *8* | *control ent. educ.* |
| Math-spec. team | 7 | 6 | 1 | 15 | 14 | treatment ent. educ. |
| | *0* | *1* | *0* | | *1* | *control ent. educ.* |
| Verb-spec. team | 2 | 7 | 6 | 21 | 15 | treatment ent. educ. |
| | *2* | *2* | *2* | | *6* | *control ent. educ.* |
| Total | 57 | 56 | 66 | 179 | 145 | treatment ent. educ. |
| | | | | | *34* | *control ent. educ.* |

Note: The unit of observation in this table is the number of teams and only includes the teams from the final sample (i.e., for which I received the information on team outcomes). The average team size is 5,76. The total sample of individuals participating in the evaluation of the entrepreneurship education program described in Chapter 3 consists of 1159 (328) children in the treatment (control) group in the school year of 2009-2010 and 842 (422) children for the treatment (control) group in the school year 2010-2011.

Table A2.2: Correlation matrix entrepreneurship education evaluation, team types and sustainability treatments

| | baseline | non-incentivized CSR | incentivized CSR | treatment entrepr. education |
|---|---|---|---|---|
| JAT-team | -0.10 | -0.04 | 0.13* | 0.02 |
| | (0.14) | (0.60) | (0.05) | (0.81) |
| Mixed-specialist team | 0.08 | -0.07 | -0.01 | -0.03 |
| | (0.22) | (0.27) | (0.88) | (0.69) |
| Math-specialist team | 0.07 | 0.09 | -0.16** | 0.08 |
| | (0.26) | (0.18) | (0.02) | (0.22) |
| Verb-specialist team | -0.10 | 0.08 | 0.02 | -0.08 |
| | (0.14) | (0.23) | (0.78) | (0.23) |
| Treatment entrepr. | -0.02 | 0.05 | -0.03 | |
| education | (0.74) | (0.43) | 0.67 | |

Note: This table reports the pairwise correlation coefficients between the treatments, p-values in parentheses. */** indicates significance at the 10%/5%-level.

# Chapter 3

# Early entrepreneurship education

*This chapter is based on Huber, L.R., R. Sloof, and M. van Praag (2014), The Effect of Early Entrepreneurship Education: Evidence from a Field Experiment, European Economic Review, 72:76-97.*

## 3.1 Introduction

Can entrepreneurship be taught? This question has been the subject of discussion for many years (e.g., Lindquist et al., 2013; Colombier and Masclet, 2008). The sharp increase in the number of entrepreneurship education programs suggests that the general consensus is that entrepreneurship can indeed be taught. From a policy perspective this is an appealing thought. The idea that entrepreneurs are not necessarily born but can also be developed creates a window of opportunity for (educational) policies aimed at enhancing entrepreneurship. However, there is little research on the effectiveness of such educational programs.

In this study we evaluate the effectiveness of an early entrepreneurship education program. A theoretical motivation to look at *early* entrepreneurship education is provided by Cunha and Heckman's (2007) general model of the technology of skill formation. This model emphasizes the importance of early investments in both cognitive and non-cognitive skills. It strongly suggests that an investment in skills not only has a direct impact on the current stock of skills but also produces spill-over effects in subsequent periods by boosting current skills and by making investments later in life more productive.[1] Early investments in skills may thus be particularly effective in the long run.

[1]Estimating the model using the Children of the National Longitudinal Survey of Youth from 1979, Cunha and Heckman (2008) and Cunha et al. (2010) indeed find evidence for these dynamic spill-over effects.

Obviously, the (potential) future spill-over benefits of early investments in skills only occur if the early investment has an immediate impact on the stock of skills in the first place. In this paper we therefore evaluate the *direct* (short term) effect of early entrepreneurship education. We report the results from a randomized field experiment using *BizWorld*, one of the leading, internationally renowned entrepreneurship education programs for primary schools. BizWorld aims to teach children aged 11 or 12 the basics of business and entrepreneurship and to promote teamwork and leadership in the classroom through an experiential learning program that takes five days (within a time span of 2 to 4 weeks). Based on the mission of BizWorld and entrepreneurship education policies more generally, we measure the effect of the program on the development of entrepreneurship knowledge, non-cognitive entrepreneurial skills and entrepreneurial intentions. The sample consists of 63 different primary schools (118 classes, 2,751 pupils) in the western part of the Netherlands that voluntarily signed up for the BizWorld program in 2010 and/or 2011. We were able to randomly assign these schools and classes to either the treatment or the control group. In both treatment and control a pre-test-post-test design was used, allowing for an (unbiased) difference-in-differences estimate of the net treatment effect.

This paper's contribution is due to three main characteristics of the study. First, to the best of our knowledge, this is the only study to evaluate the effects of entrepreneurship education on children in primary school (ages 11 and 12). Previous studies of the impact of youth entrepreneurship education follow adolescents. Second, unlike previous studies, we study the development of both knowledge and skills. Finally, we are able to estimate the unbiased (short term) effect of early entrepreneurship education on knowledge and skill development by conducting a randomized field experiment.

To evaluate the effect of the BizWorld program we selected nine non-cognitive skills from the literature that are known to be associated with entrepreneurial choice and/or success.[2] The results indicate that the program has a significantly positive effect on these non-cognitive entrepreneurial skills. On average, the skill levels in the treatment group increase to a larger extent than in the control group for all nine skills tested. The results are significant for seven skills. Self-reported scores on (constructs of) *Risk taking propensity*, *Creativity*, *Need for Achievement*, *Self-Efficacy, Pro-activity, Persistence* and *Analyzing* all increase significantly more in the treatment group than in the control group. These non-cognitive skills are not only relevant within an entrepreneurial context. There is an emerging body of research that emphasizes the importance of non-cognitive skills in predicting future labor market outcomes (Heckman et al., 2006; Cunha and Heckman, 2008; Heckman et al., 2013). For example, in

[2] An overview of the skills and their association with entrepreneurial choice and performance will be provided in Section 3.2.3.

the Perry Pre-school program it was not an increased IQ but rather the increase in non-cognitive skills that caused the difference in labor market outcomes between the treatment and the control group years later (Heckman, 2006). Moreover, the improvements in labor market outcomes reported by Chetty et al. (2011) as a result of the project STAR were caused by improvements in personality skills and behavior, rather than by increased test scores. Hence, entrepreneurship education could not only be beneficial to enhance successful entrepreneurship, but also to positively affect labor market outcomes in general. We find that the program is less effective in developing entrepreneurship knowledge. That is, there is no significant impact of the program on this outcome. Furthermore, the results indicate that, if anything, the program has a negative effect on the entrepreneurial intentions of children.

We note that the results reported here reflect the total treatment effect. Possibly, these effects of the program are not (entirely) related to the entrepreneurship component of the program. The fact that children work together in a team in a competitive environment is quite different from the regular school setting. We provide some descriptive evidence that part of the treatment effect could be driven by the teamwork component of the program. However, due to the current set-up of our field experiment we are unable to investigate the effects of the different components of the program separately.

The findings presented above, especially on non-cognitive skill development, are quite different from the mixed results found in the impact evaluation studies conducted so far (e.g. Peterman and Kennedy, 2003; Souitaris et al., 2007; Oosterbeek et al., 2010; von Graevenitz et al., 2010). All of these studies measure the effectiveness of entrepreneurship programs aimed at adolescents in secondary or higher education and most of them focus on the impact on entrepreneurial intentions only. Some studies find positive effects on entrepreneurial intentions (Peterman and Kennedy, 2003; Souitaris et al., 2007), while others find no or even a negative effect (Oosterbeek et al., 2010; von Graevenitz et al., 2010). Part of the explanation for the mixed findings might be that the two studies finding a positive effect are based on non-random assignment; self-selection may then lead to an upwardly biased estimate of the program's impact. Only Oosterbeek et al. (2010) measure the impact on the development of entrepreneurial skills, besides intentions. They find insignificant effects for a student mini-company program that is part of the international 'Young Enterprise' program offered by the Junior Achievement Worldwide network.[3]

[3]Recent studies by Karlan and Valdivia (2011) and Fairlie et al. (2012), using randomized experimental designs, report mixed results on the impact of entrepreneurship training for entrepreneurs. Karlan and Valdivia (2011) find positive effects on business knowledge. However, neither of the studies finds an (positive) impact of entrepreneurship training on business outcomes (also see McKenzie and Woodruff (2014) for an extensive overview of Business Training and Entrepreneurship evaluations).

Compared to the results found by Oosterbeek et al. (2010), our results tentatively suggest that it might be more efficient to invest in the development of entrepreneurial skills of children rather than of adolescents. On top of the large immediate (short term) impact that we measure, the empirical literature on the technology of skill formation inspired by Cunha and Heckman (2007) suggests that early investments may also have positive spill-over effects to later periods.[4]

The remainder of this chapter is structured as follows. In Section 3.2 the research design is described. Section 3.3 reports the empirical findings. Section 3.4 discusses some potential driving mechanisms underlying our treatment effect and concludes.

## 3.2 Data and methodology

### 3.2.1 Design of the field experiment

To estimate the impact of BizWorld on the development of pupils' knowledge, non-cognitive skills and intentions, a randomized field experiment was conducted between February and July in 2010, and again during the same period in 2011. In January of both years the BizWorld foundation provided us with a list of Dutch schools that planned on participating in the program next spring. In total, 120 schools signed up in 2010 (58 in the western part of the country) and 153 schools in 2011 (55 in the western part). To be able to monitor each school closely, we focus on schools close to Amsterdam. This is where our University is located, in the densely populated western part of the country (where 37% of the population lives).

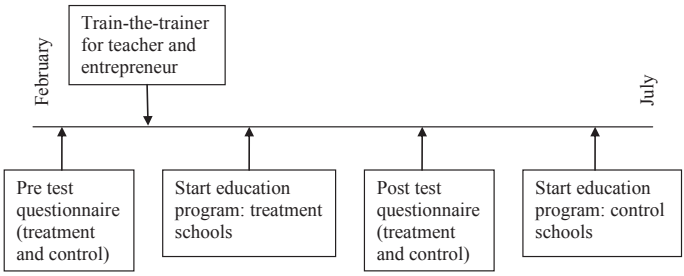
Due to the endogeneity of the participation choice at the school level, it is not possible to compare schools that chose to participate with schools that did not sign up for the program. Therefore, the schools or classes in the treatment group *and* in the control group were randomly selected from the group of schools that signed up for the program. Thus we assure that all schools in our sample have the same predisposition towards entrepreneurship (education).[5] Random assignment to the treatment or control group takes place at the class level. Hence, for schools with more than one class in the final grade it is possible that one class was assigned to the treatment group and the other class to the control group.[6]

[4]Our study is not directly comparable to Cunha and Heckman (2008), Cunha et al. (2010) and Heckman et al. (2013). We focus on knowledge and non-cognitive skills specifically related to entrepreneurship (see Oosterbeek et al., 2010) and they focus on a more general set of cognitive and non-cognitive skills (see Section 3.2.3). Moreover, BizWorld is a much smaller intervention than the Perry Preschool program (Heckman et al., 2013). However, the results we find are consistent.

[5]This means that if there is self-selection with respect to the participation in the program, it is only at the school level. This can, at most, affect the external validity of our results, not the internal validity.

[6]Overall there are eight schools in the sample where, within one year, one of the classes was part

**Figure 3.1.** Time line field experiment




We used a wait-listed control group approach, i.e., classes assigned to the control group were not excluded from participating in the education program. We merely exploited the fact that the period in which the lessons were to be conducted was flexible (i.e., somewhere between March and July). After we had completed the random assignment, the actual dates for the program were determined by mutual agreement between the teacher and the entrepreneur. In the classes in the control group the program was taught a month or two later than in the classes in the treatment group, to make sure that the treatment group has completed the program in the meantime and leaving enough time for the control group to run both the pre and post measurement (see below). The timing of the field experiment is shown in Figure 3.1.

To gather the required information for determining the effect of the education program, all pupils had to complete two extensive questionnaires, measuring not only knowledge, skills and intentions but also a wide array of individual background characteristics (see Appendix B). The first questionnaire, accompanied by a letter including some information for the parents about the research project, was sent out to all schools in the sample at the same time (in February of both years).[7] Schools were demanded to have their pupils fill out the questionnaire as soon as possible and we explained to those schools in the control group the purpose and importance of a control group in this type of research.

During the train-the-trainer session for teachers and entrepreneurs prior to the program, the details of the research project were extensively explained and discussed. Moreover, it was emphasized that the teachers and entrepreneurs should not deviate from the course content described in the instruction manual. We visited (the teacher of) every school after they had finished the education program to check their compliance with the course guidelines and to encourage response to the second questionnaire.

of the treatment group and another class was part of the control group.

[7]In the communication towards the parents, the teachers and the entrepreneurs only general information about the research project was given, no details about the evaluation procedure or measures were conveyed.

The second questionnaire was sent out to both treatment and control schools leaving approximately the same time span between the two questionnaires for both groups. For the control group we emphasized that the questionnaires had to be completed before the start of the education program, i.e., before the first introductory lesson. The pupils of the treatment schools were asked to fill out the second questionnaire after the program. Both questionnaires were carried out under the supervision of the class teacher. The lay-out of the questionnaires was specifically designed in such a way that the responses could be scanned and coded by a specialized computer system. To ensure the objective measurement of all the outcome variables, the responses from the entire survey were evaluated only by the researchers (not by the teacher or the entrepreneur).

This research design has some drawbacks. Most prominently, we cannot measure long term treatment effects due to the fact that all children in our sample eventually participate in the program. However, establishing *direct* short term effects provides a (necessary) first step in the investigation if the model of skill formation (as proposed by Cunha and Heckman (2007)) also holds for the development of non-cognitive entrepreneurial skills and knowledge. Furthermore, the current set-up of our field experiment allows us to estimate the overall treatment effect of participating in this entrepreneurship education program. However, it prevents us from estimating the influence of the different components of the program separately (e.g., learning about entrepreneurship, working in a team or being taught by an entrepreneur).

### 3.2.2 Sample

All schools that signed up for BizWorld in the western part of the Netherlands, i.e., 58 and 55 in 2010 and 2011 respectively, were contacted by the beginning of February in the respective years. We informed them about and invited them to participate in the research project. In total, of the 58 (55) schools in our research population 12 (16) schools refused participation in 2010 (2011).[8] Our resulting sample consists of $46 + 39 = 85$ schools consisting of $64 + 54 = 118$ classes and 2,751 pupils in the last grade ($2010 + 2011$).[9] Because the program is executed at the class level, we treat classes as the unit of observation, not schools.[10]

[8] In 2010 (2011), 3 (4) had objections against the research project and 9 (6) schools eventually decided to drop out of the education program. In 2011 another 6 schools were disqualified from the sample because they had already started the education program before we could send them the first questionnaire.

[9] At the school level there was an overlap between 2010 and 2011 resulting in a sample of 63 different schools.

[10] A robustness check will be shown that confirms the validity of this practice. The validity check will address the possible effects of assuming independence of observations at the class level (i) of multiple class observations within one school in the same year and (ii) within schools that participated twice (2010 and 2011). Appendix Table A3.1 shows the distribution of schools in the sample with one, two and more classes that participated in one or both years in the program.

**Table 3.1.** Sample composition

| | classes | | pupils | |
|---|---|---|---|---|
| | initial assignment | final participation | Full sample | Final sample |
| Treatment | 77 | 85 | 2001 | 1729 |
| Control | 41 | 33 | 750 | 684 |
| Total | 118 | 118 | 2751 | 2413 |

Table 3.1 shows the sample composition. 77 classes have been randomly assigned to the treatment group and 41 classes to the control group (Column 1).[11] However, some classes had to be switched from the control group to the treatment group or the other way around after the initial assignment (but before the start of the program). Teachers and entrepreneurs often met for the first time at the train-the-trainer session and planned the dates for the program there. Sometimes, their joint calendars didn't allow participation in the assigned control group (21 classes) or treatment group (13 classes).[12] The second column of Table 3.1 shows the realized sizes of the treatment (85 classes) and the control group (33 classes), whereas the right hand side of the table (Column 3 and 4) shows the distribution of *pupils* over the treatment and control groups (1,729 versus 684 in the final sample). The full sample consists of 2,751 pupils who have filled out at least one of the two questionnaires, whereas the final sample only includes those pupils who have filled out both questionnaires (n=2,413).[13] The overall response rate is $87,7\%$. Because we are interested in the development of individuals over time, our final sample consists only of the observations of those children for whom we have received both questionnaires.

### Internal Validity

An important assumption underlying the validity of the (difference-in-difference) estimation is the random assignment to the treatment and control group. In theory,

[11]The unbalancedness in the treatment assignment is related to treatment variations that we executed within the context of the program and parallel to our impact evaluation study. These other treatments pertain to variations in reward structure and in team composition (see Chapter 4 and Chapter 5). We performed additional checks (see Table A3.3 and A3.4), in which we included additional dummy variables for each treatment dimension to make sure that these treatments did not interfere with the estimation of the main results in this paper. This check confirmed that there is no systematic correlation between the development of the outcome variables and the other treatments.

[12]For participation in the control group the program should be planned later in the Spring such that the second questionnaire could be filled out before the start of the program. On the contrary, for participation in the treatment group the program should be run sufficiently early in the Spring semester leaving enough time between the end of the program and the summer holidays to complete the second questionnaire.

[13]In 2010 all classes returned the pre-treatment questionnaires and only one class did not fill out the second questionnaire. In 2011 the first questionnaire was missing for one class, and the second for four classes. Some questionnaires were missing in both years due to the absenteeism of individual children at 'test' days.

our procedure should have resulted in random assignment of children with different (observed and unobserved) characteristics to the two groups in the sample at $t = 0$. However, two changes that occurred between the initial random treatment assignment and the final treatment participation (see Table 3.1) possibly contaminate the research design: (i) The reshuffling of classes between the treatment and control group after the initial assignment and (ii) possibly selective attrition from the sample between the pre- and post-measurements.

A comparison between the observed characteristics of the individuals in the treatment and control groups in the final sample shows hardly any differences in the pre-treatment outcome variables and background characteristics, see Table 3.2 Columns 9 to 11.[14]

To address the potential problem of non-random reshuffling of classes from the treatment to the control group or vice versa after the initial assignment, we will re-estimate the main specification while removing the classes that switched between treatment and control group from the sample. Section 3.3.3 will show that the results from this estimation are almost identical to our main results. Furthermore, to alleviate concerns regarding non-random attrition, Table 3.2 shows that the differences between the treatment and the control group are very similar in the full and the final samples. In addition, separate regressions per outcome variable also show that attrition is random.[15]

Finally, we also checked with the teachers whether the children in the control group were systematically engaged in activities specifically aimed at changing entrepreneurial skills and intentions at the time of our field experiment. We acknowledge that this would be unlikely, especially given the fact that they intend to participate in the treatment program a bit later. Indeed, the check confirms that this is not the case.[16]

We conclude that there are no observed pre-treatment differences between the treatment and control group. Hence, the random assignment was not contaminated by the reshuffling of classes from the treatment to the control group after the initial assignment. Additional checks confirmed that there is no selective attrition. Together, these

[14]The only significant difference is that a larger part of the children attending Roman Catholic schools is part of the treatment group, whereas a larger part of the children attending Protestant schools has been (accidentally) assigned to the control group. The percentage of children attending public schools, however, is the same for both groups. We compared the (observed) individual characteristics of the children going to Roman-Catholic and Protestant schools and we found no significant pre-treatment differences between these two groups.

[15]In these regressions the dependent variable is an indicator for whether or not the outcome variable is observed, and the explanatory variable is the treatment dummy.

[16]There were two exceptions: in 2010 one school participated in a micro-finance program in the month prior to the entrepreneurship education program (i.e., at the time the pre-test was completed). In 2011 another school was part of an entrepreneurial primary school project (not specifically designed for the children in the last grade). Estimating the treatment effect without these schools confirmed that the results remain the same.

results show that the estimated treatment effect is indeed causal. Furthermore, we are confident that the measured treatment effects are not biased (downwards) due to the engagement in the same kind of program by classes in the control sample.

**External Validity**

The external validity of this experiment could be limited for two reasons. First, the program might be a-typical in this sample due to the research project. Second, the sample itself might not be representative for the population studied. With respect to the program there is little that can be tested. However, the large number of schools involved in the project and our small influence on the execution practice makes us confident that the program tested is very similar to the general practice in The Netherlands. We acknowledge, though, that the program is slightly different in The Netherlands from elsewhere, for instance in the United States, where the involvement of entrepreneurs is lacking.

Concerning the representativeness of the sample (for the Dutch population of school kids in the last grade of primary school) we test whether there are statistical differences between the sample and the population in terms of individual, school and neighborhood characteristics.[17] The information on those characteristics was collected by means of the questionnaires, schools' websites and Statistics Netherlands, respectively.

The pre-treatment individual background characteristics for the entire sample are shown in the first column of Table 3.2. As expected, girls make up 50% of our sample and the average age is 11.5 years. The distribution of the intended future high school track - its measure based upon the pupils' (self-reported) registration in these tracks for the next school year - is also in accordance with the national distribution.[18] Approximately 8% of the mothers of the children in the sample is an entrepreneur and 16% of the fathers run their own business, which is also in line with the countrywide average of 11% and 18% percent among working mothers and fathers, respectively. The percentage of children in the sample with a Dutch background (i.e., whose parents are both born in the Netherlands) is 56%, and somewhat lower than for the population (79%). The fraction of Surinam, Turkish and Moroccan children in our sample is higher, i.e., 8.8%, 3.5% and 4.1% respectively compared to approximately 2% for each of these in the population. This difference could be caused by the limitation of our population to the large urban areas in the western part of the Netherlands, where the ethnic diversity is largest.

At the school level, the sample seems fairly representative too. The average class

[17] Each neighborhood is characterized by a four-digit postal code (see www.cbsinuwbuurt.nl).

[18] The high school tracks in the Netherlands range from pre-vocational secondary education (1) over senior general secondary education (3) to pre-university education (5), with combination tracks in between.

**Table 3.2.** Pre-treatment differences between the treatment and the control group

| | **Total sample** | **Full sample** | | | | | **Final sample** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Treatm. + Contr. | Control | Treatment | diff (T-C) | SE | p-value | Control | Treatment | diff (T-C) | SE | p-value |
| **Non-cognitive entrepreneurial skills:** | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| Risk Taking | 4.415 | 4.386 | 4.425 | 0.0387 | (0.075) | 0.605 | 4.368 | 4.417 | 0.0491 | (0.076) | 0.522 |
| Creativity | 4.361 | 4.303 | 4.382 | 0.0786 | (0.079) | 0.323 | 4.293 | 4.385 | 0.0919 | (0.081) | 0.260 |
| Need for Achievement | 4.542 | 4.545 | 4.541 | -0.0040 | (0.082) | 0.962 | 4.542 | 4.540 | -0.0017 | (0.085) | 0.984 |
| Self-Efficacy | 4.157 | 4.151 | 4.159 | 0.0077 | (0.064) | 0.904 | 4.137 | 4.164 | 0.0272 | (0.066) | 0.682 |
| Social Orientation | 5.041 | 4.994 | 5.058 | 0.0644 | (0.072) | 0.372 | 4.986 | 5.063 | 0.0765 | (0.075) | 0.309 |
| Pro-activity | 4.574 | 4.558 | 4.580 | 0.0212 | (0.067) | 0.751 | 4.562 | 4.572 | 0.0094 | (0.067) | 0.889 |
| Persistence | 4.899 | 4.937 | 4.884 | -0.0527 | (0.065) | 0.423 | 4.934 | 4.886 | -0.0480 | (0.068) | 0.484 |
| Analyzing | 4.219 | 4.202 | 4.225 | 0.0221 | (0.068) | 0.747 | 4.193 | 4.228 | 0.0349 | (0.073) | 0.632 |
| Motivating | 4.856 | 4.846 | 4.860 | 0.0139 | (0.064) | 0.827 | 4.834 | 4.863 | 0.0297 | (0.069) | 0.666 |
| **Intentions and knowledge:** | | | | | | | | | | | |
| % Entrepreneur intentions | 0.249 | 0.267 | 0.243 | -0.0238 | (0.021) | 0.260 | 0.276 | 0.250 | -0.0265 | (0.022) | 0.238 |
| Own business (0: no, 1: maybe, 2:yes) | 1.126 | 1.098 | 1.137 | 0.039 | (0.033) | 0.237 | 1.083 | 1.140 | 0.056 | (0.034) | 0.098* |
| % Entrepreneurship knowledge | 0.729 | 0.729 | 0.729 | -0.0003 | (0.019) | 0.987 | 0.755 | 0.746 | -0.0088 | (0.017) | 0.613 |
| **Background (individual)** | | | | | | | | | | | |
| % female | 0.503 | 0.497 | 0.506 | 0.0091 | (0.020) | 0.650 | 0.499 | 0.510 | 0.0113 | (0.020) | 0.577 |
| Age pre-test | 11.625 | 11.650 | 11.616 | -0.0343 | (0.033) | 0.305 | 11.642 | 11.610 | -0.0317 | (0.035) | 0.371 |
| High school track (1: pre-voc - 5: pre-uni) | 2.946 | 2.930 | 2.952 | 0.0219 | (0.133) | 0.869 | 2.951 | 2.952 | 0.0013 | (0.136) | 0.993 |
| Nationality parents: Both non-dutch | 0.322 | 0.360 | 0.308 | -0.0522 | (0.062) | 0.405 | 0.336 | 0.280 | -0.0563 | (0.064) | 0.380 |
| % mother entrepreneur | 0.075 | 0.076 | 0.074 | -0.0015 | (0.014) | 0.910 | 0.082 | 0.076 | -0.0061 | (0.015) | 0.675 |
| % father entrepreneur | 0.152 | 0.136 | 0.158 | 0.0219 | (0.021) | 0.297 | 0.146 | 0.165 | 0.0186 | (0.022) | 0.406 |
| Education mother (1: uni - 4: no high school) | 2.059 | 2.015 | 2.076 | 0.0605 | (0.097) | 0.535 | 2.029 | 2.074 | 0.0454 | (0.099) | 0.649 |
| Education father (1: uni - 4: no high school) | 1.923 | 1.930 | 1.929 | -0.0013 | (0.078) | 0.987 | 1.913 | 1.927 | 0.0200 | (0.082) | 0.809 |
| **Number of observations** | 2751 | 750 | 2001 | | | | 684 | 1729 | | | |
| **Background (school)** | | | | | | | | | | | |
| Class size | 24.25 | 23.82 | 24.41 | 0.5968 | (1.076) | 0.581 | 24.04 | 24.44 | 0.403 | (1.053) | 0.703 |
| Roman Catholic | 0.285 | 0.113 | 0.350 | 0.2365 | (0.089) | 0.01*** | 0.117 | 0.359 | 0.242 | (0.091) | 0.01*** |
| Protestant | 0.366 | 0.495 | 0.318 | -0.1763 | (0.121) | 0.150 | 0.515 | 0.302 | -0.213 | (0.120) | 0.081* |
| Public | 0.282 | 0.282 | 0.281 | -0.0013 | (0.095) | 0.989 | 0.265 | 0.290 | 0.025 | (0.095) | 0.793 |
| Other religion (Islam, Hindu) | 0.056 | 0.091 | 0.043 | -0.0472 | (0.037) | 0.201 | 0.099 | 0.050 | -0.049 | (0.040) | 0.227 |
| *Neighborhood characteristics (based on 4-digit postal code)* | | | | | | | | | | | |
| Average income per year in € | 20148 | 19848 | 20260 | 411.97 | (943.29) | 0.664 | 19985 | 20212 | 226.41 | (958.34) | 0.814 |
| **Number of observations** | 63 | 26 | 51 | | | | 25 | 48 | | | |

Note: */**/*** indicates significance at the 10%/5%/1%-level. Observations on individual characteristics are clustered at the class level. School and neighborhood characteristics are clustered at the school level.

Robust standard errors in parentheses.

size is 24 children (national average is 23.4). The distribution across (religious) denominations of the schools is also representative; 29% of the children in the sample go to Roman-Catholic schools, 37% go to Protestant schools and 28% go to public schools.[19] The school's neighborhood level statistics on income imply that the schools participating in the program are situated in a representative cross section of neighborhoods.[20]

### 3.2.3 Outcome variables

Based on the mission of BizWorld and entrepreneurship education policies, we measure the development of the following individual outcome measures: non-cognitive entrepreneurial skills, entrepreneurship knowledge, and intentions to become an entrepreneur.

*Non-cognitive entrepreneurial skills*

Since the early sixties, entrepreneurship researchers have been interested in which non-cognitive skills are associated with (successful) entrepreneurship (see for instance Begley and Boyd (1987), Sexton and Bowman (1985) and Hornaday and Aboud (1971)). Following the study by Oosterbeek et al. (2010) we selected nine non-cognitive skills from the literature that are known to be associated with entrepreneurial choice and/or success and that, moreover, can be measured in a valid way in the realm of the current field experiment among pupils of 11 or12 years old. These relationships are summarized in Table 3.3.[21]

Ever since Knight (1921) *risk taking propensity* has been defined as one of the distinguishing characteristics of entrepreneurs (see for example Kihlstrom and Laffont, 1979; Kanbur, 1979). Subsequent empirical research has mostly shown that entrepreneurs have a lower degree of risk aversion than others (Stewart and Roth, 2001; Cramer et al., 2002; Hvide and Panos, 2013).[22] As already noted by Schumpeter (1934) entrepreneurs must be able to generate new ideas and form new combinations, i.e. to be successful as an entrepreneur a person must be *creative*. Another characteristic that is traditionally associated with entrepreneurship is *need for achievement* (McClelland, 1965; Shane and Venkataraman, 2000). That is, an entrepreneur sets challenging goals and continuously

[19]Note that (almost) all primary schools in the Netherlands, irrespective of their denomination, are publicly funded, i.e., there is a 'money follows pupil' system.

[20]The average gross income in these neighborhoods is €20.147 per income recipient per year, whereas the national average is €24.100 for couples with children below the age of 18 and €16.100 for single parents with children below the age of 18.

[21]We note that the empirical evidence on the association between many of these skills and entrepreneurship is not conclusive, thus what we report are commonly found associations (see e.g. Zhao et al. (2010) and Parker (2009) for an overview).

[22]However, the empirical evidence is rather mixed, including contradicting (e.g. Brockhaus, 1980), non-linear (Caliendo et al., 2010) and insignificant results (e.g. Parker, 2008).

**Table 3.3.** Non-cognitive entrepreneurial skills and knowledge

| Outcome variables | Definition | association with entrepreneurial *choice* | association with entrepreneurial *success* | Cronbach's $\alpha$ |
|---|---|---|---|---|
| *Non-cognitive skills* | | | | |
| Risk taking | Predisposition towards risky alternatives | + | ∩ | 0.75 |
| Creativity | Ability to create many opportunities | + | + | 0.75 |
| Need for achievement | Desire to do well | + | + | 0.69 |
| Self-efficacy | Belief in own ability | + | + | 0.67 |
| Social orientation | Ability to make useful connections | + | + | 0.63 |
| Pro-activity | Willingness to take action | + | + | 0.58 |
| Persistence | Ability to continue despite setbacks | + | + | 0.61 |
| Analyzing | Ability to assess complex situations | 0 | + | 0.56 |
| Motivating | Ability to inspire or stimulate subordinates | 0 | + | 0.80 |
| *Entrepreneurship knowledge* | Knowledge about running a business | + | + | |

Note: A '+' indicates that the existing literature has established a positive relationship between the skill and entrepreneurial intentions or success , '0' indicates no association has been established and '∩' refers to an association that follows an inverse U-shape.

seeks to improve his or her performance (Begley and Boyd, 1987). Furthermore, Chen et al. (1998) find that *self-efficacy* is positively associated with probability of becoming an entrepreneur, because confidence in one's own ability increases the willingness to pursue entrepreneurial opportunities. Moreover, several empirical studies have shown that *social orientation* is important for becoming an entrepreneur as well as for the success rate of new ventures (e.g. Davidsson and Honig, 2003; Dahl and Sorenson, 2012; Roberts and Sterling, 2012). Social orientation is the ability to benefit from social connections and from interactions with others (Glaeser et al., 2002). The relationship between *pro-activity* and *persistence* and entrepreneurship has also been studied and is found be positive for the start-up of a company and for subsequent venture growth (Baum et al., 2001). *Analyzing* refers to analytical or problem solving skills. It is the ability to create or spot opportunities by systematically analyzing and solving a problem, and is thus a relevant skill for entrepreneurs (Ward, 2004; Baron and Ensley, 2006). Finally, *motivating* skills are associated with new venture growth (Baum and Locke, 2004) as well as better labor market outcomes in general (Borghans et al., 2006). Table 3.3 provides an overview of the relationships established in the literature between the non-cognitive skills we measure and entrepreneurial choice (Column 3) and success (Column 4).

The separate skills presented in Table 3.3 are not solely important for entrepreneurs, but are powerful predictors of social economic success in general (e.g. Heckman et al., 2006, 2013). Moreover, any direct effect could induce future spill-over effects (through dynamic complementarity and self-productivity of skills) and thereby make early in-

vestments in non-cognitive skills even more effective in the long run.[23] However, the non-cognitive skills used in our study are not directly comparable to those studied by Cunha and Heckman (2008), Cunha et al. (2010) and Heckman et al. (2013). The non-cognitive skills they study are measured through the Behavior Problem Index in the first two papers, and the Pupil Behavior Inventory (PBI) in the latter. Both tests measure childhood temperament traits and precede the well known (and commonly used) Big Five traits of personality inventory. Certain aspects of the Big Five personality traits, i.e., Conscientiousness, Openness to Experience and Emotional Stability have been positively associated with entrepreneurial choice and success (Zhao et al., 2010). The non-cognitive skills we use are related to these Big Five traits, e.g. *persistence* and *need for achievement* are related to Conscientiousness, and *creativity* and *pro-activity* are related to Openness to Experience (see Almlund et al. (2011, Table 3) for a complete overview of the Big Five traits and their facets). Thus, developing these skills, separately or some combination, is beneficial for both future entrepreneurs and employees.

The non-cognitive skills are measured by means of a validated self-assessment test. Self-reported paper and pencil tests are the most widely used measures in personal psychology (Borghans et al., 2008). Recent psychological studies have confirmed the validity of the use of self-assessment tests in middle and late childhood, i.e., for children between 8 and 12 years old (Barbaranelli et al., 2003; McCrae et al., 2002). The test is based on the one used and further validated by Oosterbeek et al. (2010) and Hoogendoorn et al. (2013). Of course, because our study pertains to children at the age of 11 or 12 instead of (young) adults, we have developed and validated a slightly adapted version of this test. We did so in close collaboration with a child psychologist. Three elements characterize the transformation for the younger target group. First, the questionnaire is shorter than the original, using three instead of four items per skill, thus matching the concentration time span of children.[24] Second, certain constructs, such as market awareness, networking skills, etc., were excluded because they are difficult to relate to as a child. Third, we rephrased the original statements to make them easier

[23] Pfeiffer and Reuss (2008) use a simulation model calibrated to German data to get an idea of the financial returns to investments in skills that the Cunha and Heckman (2007) model may imply. Consistent with the predictions by Knudsen et al. (2006) and Borghans et al. (2008), self-productivity and direct complementarity are assumed to differ between cognitive and non-cognitive skills. In early childhood these are higher for cognitive than for non-cognitive skills, but from late childhood (10 to 11 years old) onwards this is the other way around. As a result, investments in cognitive skills are relatively more important during the pre-school years, whereas the school years play an important role in the development of non-cognitive skills. Because the positive complementarities decrease over time, the analysis of Pfeiffer and Reuss (2008) also suggests that additional investments in pre-school and primary school yield higher returns than investment impulses in secondary or tertiary education.

[24] The overall score for each skill is calculated by the weighted average of the three items. The weighting is determined by the contribution of each item to the construct based on the values calculated by a principal component analysis for each construct.

for children to understand (see also Barbaranelli et al., 2003). Examples of statements are: "I can encourage other children to do their best" (motivating), "I want to perform better than others" (need for achievement), "I like to take chances" (risk taking), and "I think I'm good at solving problems" (self-efficacy). Statements had to be answered on a seven-point scale, expressing the extent to which a child agrees with each statement (see Appendix B for the entire questionnaire).

We use (standardized) Cronbach's $\alpha$ to measure the internal consistency and validity of our measures.[25] The Cronbach's $\alpha$'s range from 0.56 to 0.80 (see the last column of Table 3.3).[26] Traditionally in the literature a cut-off of 0.70 is considered satisfactory. However, alpha is a function of the number of items in a scale (Cortina, 1993). Since we have only three items per scale, we decided to use a slightly less stringent criterion of 0.60 as a cut-off. Because the reliability of $\alpha$ as a measure of internal validity has been subjected to debate (Revelle and Zinbarg, 2009), we also conducted a principal component analysis to check the independence of the scales. This test revealed that *self-efficacy*, *need for achievement* and *pro-activity* do not load into separate factors, despite the high Cronbach's $\alpha$ for the first two constructs.[27]

The outcome variable is defined as the development in each non-cognitive skill ($\Delta y$). The development is measured per individual by the change in the score of each construct between $t = 0$ and $t = 1$ (i.e., $\Delta y_i = y_{i1} - y_{i0}$).

*Entrepreneurship knowledge*

Evidence on the relationship between entrepreneurship experience and the decision to become an entrepreneur is consistently positive (Parker, 2009). According to Shane (2003) experience includes training for skills such as selling, problem solving, organizing and communicating. These are also the type of skills and knowledge that are taught during the entrepreneurship education program. Therefore, one could expect a positive (mediating) effect of the development of entrepreneurship (related) knowledge on entrepreneurial choice (see last row in Table 3.3).

The association between knowledge and entrepreneurial success also appears to be positive. In general, human capital theory states that education increases productivity and thus leads to higher income (Mincer, 1958; Becker, 1964). In the entrepreneurship literature the link between education and business performance or entrepreneurial in-

[25] When starting with the development of the test for children, we tested the (internal) validity of our adapted measures by conducting a pilot study consisting of 118 children who participated in the BizWorld program and filled out both pre-test and post-test questionnaires in the fall of 2009. One skill (*Flexibility*, $\alpha = 0.10$) was removed from the questionnaire and another skill (*Need for power*, $\alpha = 0.46$) was replaced by *Need for achievement*.

[26] The reported Cronbach's $\alpha$ is the unweighted average of the values from the pre- and post-test questionnaire. The average spread between these two measurements is 0.04.

[27] We will thus apply some caution when interpreting the results for these measures.

come has also been widely established (Bates, 1990; Robinson and Sexton, 1994; Fairlie and Robb, 2007). Moreover, the meta-analysis conducted by Unger et al. (2011) shows that there is a significant positive relationship between task-related human capital and entrepreneurial success.

One of the desired results of the BizWorld program is the development of knowledge that is relevant for entrepreneurship, i.e., knowledge about what an entrepreneur does and what it entails to run a business. A set of seven specific multiple choice questions is used to measure this knowledge. Examples are: "If a company makes less revenue by selling products or services than it spends, it will... a) be registered at the stock market, b) make a profit, c) make a loss, d) have debts", and "To set the price of a product you have to take into account... a) how much it costs to make the product, b) how many products can be made in a certain amount of time, c) the price that competitors ask for their products, d) all of the above". The outcome variable is the development of entrepreneurship knowledge, which is measured by a change, between $t = 0$ and $t = 1$, in the percentage of correct answers to these questions.[28]

*Entrepreneurial intentions*

In addition to the main outcome variables, we measure the impact of the program on the children's intentions to become an entrepreneur. Although raising entrepreneurial intentions is not a specific goal of the program, it is one of the main goals of entrepreneurship education in general and it is frequently used as an outcome measure in other impact evaluation studies. However, as mentioned in the introduction, findings on the effect of participation in an entrepreneurship education program on entrepreneurial intentions are mixed (e.g. Oosterbeek et al., 2010; Peterman and Kennedy, 2003).

The measurement of entrepreneurial intentions at the age of 12 is difficult and no precedents are available to indicate the validity or predictive power of any such measure. We use two different measures to estimate the change in the intention to start a business as a result of program participation. First, children were asked to select a maximum of three jobs they might like for their future occupation from a list of 22 professions, one of which was 'entrepreneur - (boss in your own company)'. A dummy variable (Future job: entrepreneur) is created to indicate whether *entrepreneur* was on the list of three. This was the case for a quarter of the sample pre-treatment. The change in intentions is measured by the differences in this (dummy) variable between the first

[28]To prevent children from memorizing the answers to the knowledge questions, three out of the seven questions in the first questionnaire were rephrased in the second questionnaire (see Appendix B). For example, instead of asking about making a loss (as in the example question given above), the question was: "If a company makes *more* revenue by selling products or services than it spends, it will... a) be registered at the stock market, b) make a profit, c) make a loss, d) have debts". These changes were determined prior to the start of the education program and applied to the entire sample (i.e., treatment and control group).

**Table 3.4.** Descriptive statistics of outcome variables

| | Treatment | | Control | | Treat + Control | |
|---|---|---|---|---|---|---|
| **Outcome variables** ($\Delta\bar{y} = \bar{y}_1 - \bar{y}_0$) | **Mean** | **Std. Dev.** | **Mean** | **Std. Dev.** | **Mean** | **Std. Dev.** |
| Risk Taking | 0.21 | 1.15 | 0.11 | 1.13 | 0.180 | 1.14 |
| Creativity | 0.22 | 1.17 | 0.15 | 1.10 | 0.199 | 1.15 |
| Need for Achievement | 0.25 | 1.07 | 0.08 | 1.00 | 0.197 | 1.05 |
| Self-Efficacy | 0.22 | 1.02 | 0.08 | 0.91 | 0.177 | 0.99 |
| Social Orientation | 0.11 | 1.01 | 0.07 | 0.92 | 0.098 | 0.99 |
| Pro-activity | 0.14 | 1.02 | -0.01 | 0.94 | 0.094 | 1.00 |
| Persistence | 0.03 | 1.07 | -0.10 | 1.02 | -0.009 | 1.05 |
| Analyzing | 0.22 | 1.01 | 0.11 | 0.90 | 0.190 | 0.98 |
| Motivating | 0.13 | 1.17 | 0.06 | 1.15 | 0.113 | 1.17 |
| Entrepreneurship knowledge | 0.04 | 0.21 | 0.02 | 0.21 | 0.035 | 0.21 |
| Future job: entrepreneur (0/1) | -0.003 | 0.46 | 0.02 | 0.46 | 0.004 | 0.46 |
| Own Business (0-2) | -0.09 | 0.63 | 0.07 | 0.60 | -0.047 | 0.62 |

and the second questionnaire.

The second measure of entrepreneurial intentions (Own Business) is the answer to the question: 'Do you think that you would like to start your own company one day?'; (*yes*, *no* or *maybe)*. This variable was coded in such a way that a change in the answer to this question from *no* (code 0) to *maybe* (code 1) and from *maybe* to *yes* (code 2) is regarded as a similar increase in entrepreneurial intentions. A change from *no* to *yes* is regarded as a more positive change in intentions. We will interpret the results for intentions with great care for the reasons stated before.

Table 3.4 reports the descriptive statistics of all the outcome variables for the entire sample and for the treatment and the control group separately.

## 3.3 Results

### 3.3.1 Estimation method

To analyze the effect of the BizWorld program on the outcome variables, a difference-in-differences analysis (DID) is used. The value of the outcome variable of individual $i$ in the treatment group before the start of the program ($t = 0$) is denoted by $y_{Ti,0}$, while $y_{Ti,1}$ gives the corresponding value after the treatment period ($t = 1$). For the control group, similar notation is used, i.e., $y_{Ci,0}$ and $y_{Ci,1}$. The difference between the two measures, $\Delta y_{Ti} = y_{Ti,1} - y_{Ti,0}$ and $\Delta y_{Ci} = y_{Ci,1} - y_{Ci,0}$, reports the changes in the level of each outcome variable between time $t = 0$ and $t = 1$ for an individual in the treatment or the control group respectively. The average change per outcome variable between the pre-test and the post-test of all the children in the treatment and

**Table 3.5.** Treatment effects

| | Treatment | | | | Control | | | | DID (using standardized measures) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | no controls | | with controls | |
| | (1) | (2) | (3) | | (4) | (5) | (6) | | (7) | | (8) | |
| **Outcome variables** | t = 0 | t = 1 | $\Delta y_T = y_{T1} - y_{T0}$ | | t = 0 | t = 1 | $\Delta y_C = y_{C1} - y_{C0}$ | | $\delta$ | | $\delta$ | |
| *Non-cognitive entrepreneurial skills* | | | | | | | | | | | | |
| Risk Taking | 4.41 | 4.62 | 0.208*** | (0.024) | 4.37 | 4.48 | 0.111** | (0.044) | 0.100** | (0.044) | 0.108** | (0.045) |
| Creativity | 4.38 | 4.60 | 0.216*** | (0.029) | 4.30 | 4.45 | 0.155*** | (0.043) | 0.083* | (0.045) | 0.098** | (0.047) |
| Need for Achievement | 4.54 | 4.78 | 0.245*** | (0.026) | 4.54 | 4.62 | 0.077** | (0.038) | 0.158*** | (0.049) | 0.151*** | (0.051) |
| Self-Efficacy | 4.16 | 4.37 | 0.218*** | (0.025) | 4.14 | 4.22 | 0.075** | (0.035) | 0.150*** | (0.049) | 0.156*** | (0.043) |
| Social Orientation | 5.06 | 5.17 | 0.108*** | (0.025) | 4.99 | 5.06 | 0.073** | (0.035) | 0.064 | (0.053) | 0.049 | (0.053) |
| Pro-activity | 4.57 | 4.70 | 0.136*** | (0.025) | 4.57 | 4.56 | -0.009 | (0.036) | 0.145*** | (0.051) | 0.166*** | (0.045) |
| Persistence | 4.89 | 4.91 | 0.026 | (0.026) | 4.93 | 4.84 | -0.098** | (0.039) | 0.100** | (0.047) | 0.105** | (0.046) |
| Analyzing | 4.23 | 4.45 | 0.223*** | (0.025) | 4.20 | 4.30 | 0.107*** | (0.035) | 0.130*** | (0.044) | 0.138*** | (0.049) |
| Motivating | 4.87 | 5.00 | 0.132*** | (0.029) | 4.84 | 4.90 | 0.065 | (0.045) | 0.068 | (0.047) | 0.061 | (0.048) |
| *Entrepreneurship knowledge* | 0.75 | 0.79 | 0.040*** | (0.006) | 0.75 | 0.77 | 0.019** | (0.009) | 0.071 | (0.066) | 0.042 | (0.063) |
| *Entrepreneurial intentions* | | | | | | | | | | | | |
| Future job: entrepreneur (0/1) | 0.25 | 0.25 | -0.004 | (0.011) | 0.28 | 0.30 | 0.023 | (0.018) | -0.090* | (0.050) | -0.075 | (0.050) |
| Own Business (0-2) | 1.14 | 1.05 | -0.094*** | (0.015) | 1.08 | 1.15 | 0.070*** | (0.023) | -0.213*** | (0.049) | -0.217*** | (0.051) |
| **Number of observations** | 1729 | 1729 | | | 684 | 684 | | | 2351 | | 2304 | |

Note: The estimates in each cell come from separate regressions. Observations clustered at the class level, robust standard errors in parentheses. All regressions control for the baseline level of the outcome variable. DID with controls includes individual characteristics: age, gender, future high school track, nationality parents, parents entrepreneurial status; school/neighborhood characteristics: class size, school denomination, avg. income per year and a year dummy for 2010/2011. */**/*** indicates significance at the 10%/5%/1%-level.

the control group are denoted by $\Delta y_T$ and $\Delta y_C$. Hence, the DID estimate is given by:

$$\delta = \Delta y_T - \Delta y_C \tag{3.1}$$

Double differencing removes potential biases associated with the common development of the children over time that are unrelated to the program (Imbens and Wooldridge, 2009).[29] The skills we measure, although related to entrepreneurship, can be developed in several ways. Most importantly, since all the children in the sample are in school during our observation period, some development in these skills is expected even without participation in the program. Additionally, other everyday activities, e.g. at sports clubs or other social events, could also be beneficial for the development of non-cognitive skills. Finally, part of the increase observed, in the treatment as well as the control group, might be due to the Hawthorne effect. However, assuming that any potential Hawthorne effect is equally strong in both groups (i.e., they both fill out the same questionnaires), the use of a difference-in-differences estimator will remove this potential overestimation of the treatment effect.

For each individual ($i = 1, ..., N$) the following variables are observed: $D_i, y_{i0}, y_{i1}, X_{i0}$, and $X_{i1}$. Where $D_i$ is a dummy variable which takes the value 1 if individual $i$ was part of the treatment group, $y_{it}$ is the outcome value for individual $i$ at time $t$, and $X_{it}$ is a vector of control variables for individual $i$ at time $t$. The difference, $\Delta y_i = y_{i1} - y_{i0}$, is then regressed on the treatment indicator, $D_i$, and the lagged outcome, $y_{i0}$:

$$\Delta y_i = \alpha + \delta D_i + \beta y_{i0} + \epsilon_i \tag{3.2}$$

For the ease of the interpretation and the comparison between the results of the different outcome variables we use standardized outcome and explanatory variables in our main specification. Furthermore, the baseline level of the outcome variable is included to correct for a potential ceiling effect (i.e., if your initial score or skill level is high, there is less room for improvement as a result of the treatment). The observations are clustered per class to obtain estimates with robust standard errors, accounting for the fact that the results for children in the same class are potentially correlated. To confirm the robustness of the estimated coefficients from Equation (3.2), we will also estimate the model with a vector of control variables ($X_i$) such as age, gender, parental entrepreneurial activity, etc.[30]

[29] To estimate the treatment effect, the panel structure of the data is used together with the unconfoundedness assumption given the lagged outcomes. The unconfoundedness assumption requires that conditional on a set of observed covariates (i.e., controls and outcomes), treatment assignment is essentially randomized (Imbens and Wooldridge, 2009, p.23). Given the set-up of our experimental design, we feel that it is safe to assume that this assumption holds in our sample.

[30] If the randomization was successful, in principle, a simple level regression should yield the same results as the DID. Estimating Equation (3.2) using the level of the outcome variable at $t = 1$ ($y_{i1}$) as

### 3.3.2 Main results

The results for the DID estimation of Equation (3.2) are shown in Table 3.5. The mean values for the outcome variables at $t = 0$ and $t = 1$ are shown for both the treatment (Columns 1 and 2) and the control group (Columns 4 and 5). Columns 7 and 8 of Table 3.5 show the net treatment effect, $\delta$, and the robust standard errors in parentheses.

*Non-cognitive entrepreneurial skills*

All but one of the non-cognitive entrepreneurial skills increase significantly between $t = 0$ and $t = 1$ within the treatment group. The only exception is *Persistence* for which the difference is positive, but not significant. In the control group six of the non-cognitive skills change positively and significantly in the same period. *Motivating* and *Pro-activity* do not show a significant change and *Persistence* decreases significantly for the children in the control group. The fact that the children in the control group also develop their skills in this time frame shows that they do not spend the time that the treated children spend on the program idly. They develop their non-cognitive skills through the regular lessons offered. This emphasizes the importance of a control group in our research design.

The results for the DID analysis show that the difference in development between the treatment and the control group is positive for all non-cognitive skills. The change in these outcome variables is larger in the treatment group than the control group. The treatment effect is statistically significant for seven out of the nine skills: *Risk taking propensity*, *Creativity*, *Need for Achievement*, *Self-efficacy, Pro-activity, Persistence* and *Analyzing*.[31] The last column (Column 8) of Table 3.5 shows that the treatment effects remain the same or increase slightly when we control for individual, school and neighborhood characteristics as well as the year of the data collection.[32]

The size of the treatment effects we find is substantial. For instance, children in our treatment group show a significant increase in Creativity of 0.10 of a standard deviation compared to the control group. Self efficacy, Risk taking and Need for Achievement increase by 0.16, 0.11 and 0.15, respectively. Overall, the results show that the effect sizes are between 0.05 and 0.16 of a standard deviation. Moreover, on top of this immediate (short term) impact, early investments may also induce future spill-over effects (through dynamic complementarity and self-productivity of skills) and thereby make early investments in non-cognitive skills even more effective in the long run.

the dependent variable (both with and without control variables) indeed gives the same results as the ones presented in the next section. We choose to report the DID estimates as our main specification, because we feel that the development of skills and knowledge is a more interesting outcome variable than the level (of these outcome variables) at a certain point in time.

[31] Note that *Self-efficacy, Need for achievement* and *Pro-activity* did not load into separate factors and therefore require careful interpretation of the results for these measures.

[32] See Table A3.5 in the Appendix for the detailed estimation results pertaining to the controls.

*Entrepreneurship knowledge*

The estimated effect on entrepreneurship knowledge, can also be found in Table 3.5.[33] Both within the treatment group as well as in the control group there is a significant increase in the percentage of correct answers between $t = 0$ and $t = 1$. The increase is slightly larger in the treatment than in the control group, which results in a positive, yet insignificant, estimate of the net treatment effect ($\delta$). The picture remains unchanged when we include the set of control variables. Therefore, the program does not seem to have the intended effect on the development of entrepreneurship knowledge.

*Entrepreneurial intentions*

The results for the first intention measure, i.e., future job choice, show that the intention towards becoming an entrepreneur decreases slightly within the treatment group and increases slightly within the control group between $t = 0$ and $t = 1$. This results in a negative and marginally significant estimate of the net treatment effect without controls. The result is insignificant when controlling for individual, school and neighborhood characteristics.

The results from the second measure show that the intention to start a business some time in the future decreases significantly for the children in the treatment group, whereas the children in the control group show a significant positive change in this intention. Therefore, the DID estimate for this intention measure (from both equations) is significantly negative. Thus, in line with the results found by Oosterbeek et al. (2010), we find that, if anything, this entrepreneurship education program has a negative effect on the intention towards becoming an entrepreneur. Alternatively, the program could have an indirect effect on entrepreneurial intentions. The non-cognitive entrepreneurial skills that we measure are (almost) all positively associated with entrepreneurial intentions. Hence, the significantly positive effect of the program on these skills, might (positively) influence the intention to become an entrepreneur in the future. As mentioned before, due to the lack of validated measures of entrepreneurial intentions for children, we treat these results with caution.

### 3.3.3 Robustness checks

The results from the previous section show that our findings are robust when we include a variety of individual, school and neighborhood characteristics. We perform several more robustness checks.

First, as announced, we estimate Equation (3.2) excluding the classes from the

[33] The detailed estimation results for entrepreneurship knowledge and entrepreneurial intentions are shown in Table A3.2 in the Appendix.

sample that switched between the treatment and the control group after the initial treatment assignment. A priori, the choice to switch was only guided by practical concerns and we expect no relationship with the outcome variables. Indeed, the results from these estimations are the same for most outcome variables, only for Analyzing ($\delta = 0.11$, $p - value = 0.12$) and Creativity ($\delta = 0.07$, $p - value = 0.25$) the results are slightly weaker than the main results. This may also be due to the fact that the sample size reduces from 118 to 84 classes when excluding switchers.

Second, in order to test whether the actual treatment status (i.e. the dummy for treatment participation) is in fact exogenous, we compare the estimated coefficients for this variable from the OLS with the coefficients from a 2SLS estimation (using initial treatment assignment as an instrument) using a Wu-Hausman $F$-test (Hausman, 1978) for endogeneity. For this test the null hypothesis is that the OLS estimate is consistent, i.e. that the treatment status is exogenous. We perform this estimation separately for all outcome variables (F-test range from 0.0004 to 1.29, with p-values of 0.98 and 0.26, respectively). Thus the results from this test confirm that the actual treatment participation is indeed exogenous (i.e. random).

A third robustness check indicates that it is unlikely that the results are influenced by a possible appreciation bias. For example, if the children are very enthusiastic about the program, we might be measuring the children's sheer appreciation of the program instead of actual learning. However, we measure a low positive correlation coefficient between the grade the children assigned to the education program (on a scale of 1-10) to express their appreciation of it, and their skill development, i.e., between 0.05 and 0.13.

Fourth, we rule out that the effects measured are only very short term and temporary. To this end, we measure if the impact of the time elapsed between the program and the completion of the second questionnaire on our outcome variables is negative. Time elapsed is (imperfectly) measured as the number of days between the start of the program and the day we received the second questionnaire (36 days on average, varying from 13 to 70 days, std. dev. 15 days, while the duration of the program itself was approximately two weeks on average).[34] Evidently, this test only includes the treatment and not the control group. The estimation shows that the time elapsed between the education program and the post-test questionnaire does not change our main results.[35] Additionally, we compare the the time elapsed between the receipt of the two questionnaires between the treatment and the control group, to ensure that this does not vary systematically by treatment status. The degree of variation in the timing of the responses to the pre and post-test is virtually identical in the two groups.

[34] Unfortunately this detailed information was only available for the 2011 sample.

[35] We only find a significant negative time effect on the development of Social orientation (p-value: 0.02), which was not significant in our initial estimation.

Hence, we are confident that we are indeed measuring the same developmental time trend in both groups.

Fifth, clustering observations at the school (n=63) instead of at the class level (n=118), we establish that the (significant) results remain significant. Although the children, and in some cases also the teacher, change from one school year to another, one could argue that the observations per school are potentially correlated. The results of these estimations are the same and are shown in Table A3.6 and A3.7 in the Appendix. Moreover, the randomization into treatment and control group was done at the class level. Hence, one could argue that the class, instead of the individual should be the unit of analysis. As a robustness check, we perform the same analyses at the class level and the results are very similar, albeit slightly less significant due to the loss of observations. The results for this analysis are shown in Table A3.8 in the appendix.

The findings from the checks described above show that our results are stable to various changes applied to the original specification. Therefore, we are confident that the early entrepreneurship education program we study has a robust positive effect on non-cognitive entrepreneurial skills.

### 3.3.4 Heterogeneous treatment effects

The starting point for our analysis of heterogeneous treatment effects are the control variables that have a significant impact on the outcome variables (see Tables A3.2 and A3.5 in the Appendix). For example, the development of entrepreneurship knowledge and some non-cognitive entrepreneurial skills are distinct for males and females. For all independent variables that apparently move the intercept, we test whether they are also associated with heterogeneity in effect sizes. In particular, we considered interactions with gender, age, intended high school track, school denomination, year (2010 versus 2011 or both) and the average income in the school's area. We do not find any heterogeneities for these variables. Additionally, we looked at differences between children with and without parents active as an entrepreneur. In the empirical literature there is some evidence of inter-generational transmission of entrepreneurial skills and occupational choice (Lindquist et al., 2013; Colombier and Masclet, 2008; Dunn and Holtz-Eakin, 2000). However, we do not find any significant differences in the treatment effect on any of the outcome variables for children with entrepreneurial parents. For other variables a few (insignificant) results are noteworthy.

Using the model developed by von Graevenitz et al. (2010), we test whether the change in intention was moderated by a person's entrepreneurial ability.[36] This turned

[36]von Graevenitz et al. (2010) develop a formal Bayesian updating model to explain the mixed findings on entrepreneurial intentions and predict that program participation causes a sorting effect among students with different entrepreneurial abilities. Those students who discover to be less suitable

out not to be the case: the change in entrepreneurial intentions due to treatment is the same for children with high and low pre-treatment entrepreneurial ability. We also test the proposition by von Graevenitz et al. (2010) that the decision to become an entrepreneur becomes more defined after the program, i.e., that the variance in the responses (for business ownership intentions) is larger after the program than before. However, the results do not support this proposition either. Thus, we find little evidence of sorting.

Additionally, we considered the possible effect of the size of the team on the change in outcome variables (thus excluding the control group from the sample). Most of the teams consist of five or six children, but team size can vary between four and seven members per team. Despite the greater likelihood of free riding in bigger teams, possibly leading to less active participation, we do not find smaller learning effects for larger teams, nor does team size affect entrepreneurial intentions.

All in all, because we find almost no heterogeneities in treatment effects, we conclude that the effects we establish hold by and large across the board.

## 3.4 Discussion and conclusion

### 3.4.1 Discussion

Before we reach our conclusion, we provide in this subsection an interpretation of the treatment effect that was established in Section 3.3. The results show that participation in the BizWorld education program has a robust significant positive effect on the development of non-cognitive skills. However, this entrepreneurship education program simultaneously introduces several learning aspects into the classroom that are different from the regular learning experience. First of all, the program teaches the children about entrepreneurship, which is not part of the regular (primary) school curriculum in the Netherlands. Secondly, the program involves teamwork, which can be a source of inspiration and confidence building and thereby could have a stimulating effect on the non-cognitive skills. Finally, bringing an entrepreneur or someone from the business world into the classroom to teach the course, as is done in the Dutch program, could also trigger the development of certain skills. As such, each of these (major) components could influence the development of entrepreneurial skills in its own way.

To understand which part of the program drives the overall treatment effect established in Section 3.3, we perform several (albeit imperfect) tests. For these tests we use some qualitative evidence that we collected by means of the second questionnaire,

for becoming an entrepreneur will have lower intentions after the program than those who receive positive signals during the course. They find empirical support for their sorting prediction.

i.e., after the education program.[37] To start with the entrepreneur, we look at two questions: one measures the importance of his or her role, and the other measures the children's appreciation for the entrepreneur. The appreciation is measured by the grade that the children give the entrepreneur on a scale of 1 to 10. The analysis shows that the grade is positively correlated with the development of all the non-cognitive skills. However, the correlation coefficients are small in size (between 0.05 and 0.10). Furthermore, we measure the importance of the role of the entrepreneur by looking at a question that asks the children to place the components of the program in such an order to indicate what motivates them to do their best from 1 (most important) to 7 (least important). The possible answers are: play a game, being taught by an entrepreneur, work in a team, a change to normal school days, to be able to make money, learn about business and entrepreneurship, and show what I can do. When comparing the answers "being taught by an entrepreneur" is the least important reason (rank 4.72) for the children to perform well. Hence, based on these descriptive results, the presence of the entrepreneur does not seem to have a major impact on the learning process.

The relationship between teamwork and learning is studied in various empirical papers from the economics literature as well as in education research. For instance, the paper by Hamilton et al. (2003), that studies the effect of teamwork on productivity, finds that part of the increased productivity can be attributed to mutual learning. Recent work by Hoogendoorn and Van Praag (2012) also indicates that (mutual) learning might be one of the mechanisms that explains why more ethnic diverse teams achieve better results. Research in educational settings shows that students working in small groups learn more efficiently than students solving problems individually, because teams seem to be better at handling problems with complex information (Plass et al., 2013; Kirschner et al., 2011). Moreover, several studies indicate that cooperative learning only leads to better achievements if group rewards are provided (Pai et al., 2014; Lou et al., 1996).

The optimal way to disentangle the teamwork component from the entrepreneurship part of the program would be to compare our findings to those from a very similar early entrepreneurship education program in which the whole program is done by the children individually. However, since (a study about) the counterfactual is not available, i.e., entrepreneurship education on an individual basis rather than in teams, such a specific comparison cannot be made. Therefore, we can only provide some descriptive evidence on the association between teamwork and the changes in our outcome variables. The results from these tests are presented below.

[37] Hence, the information from these questions is only available for the children in the treatment group.

First, several team characteristics, such as the mean and the variance of the initial skills and knowledge (at the team level), are added to the estimation equations. If some of these team characteristics were correlated with the learning outcomes, this would indicate that teamwork or certain team dynamics are beneficial for the skill development. None of these characteristics turn out to be important in the development of individual knowledge or non-cognitive entrepreneurial skills, nor for the changes in entrepreneurial intentions. With this test we estimate the effect of (small) differences in the team composition on the development of the outcome variables. However, since all the children in our sample work together in teams, small changes in group composition might not capture the teamwork component we are looking for.

Secondly, to shed some more light on the teamwork mechanism, we use the answers to the question "Which part of BizWorld did you like best?". The possible answers are: start-up a company, design a product, teamwork, calculations, production, sales, taking decisions, and make financial statement. The children are asked to rank the topics from 1 (like best) to 8 (like least). Seven of the possible answers are related to the business component of the program and one is about working together as a team. 6.8% of the pupils answer that they like teamwork the best, and 40.1% included teamwork in their top three of favorite parts of the program. If we look at the specific element of starting up a company, we find that for 7.1% of the children this is their favorite part and it was ranked among the top three by 35%. We also compare the overall ranking between the teamwork component and the start-up component. This comparison shows that the mean rank for teamwork is 4.01 and the start-up component has an average rank of 4.24. Both rankings are not far from the mean and the difference is small, yet significant.

Finally, we also have information on how well the team worked together. However, this measure is less precise and, as can be expected, the results show that the conditional correlation between the ex-post evaluation of how well the team worked together and the outcome variables is mostly positive.

The descriptive tests presented above show that the significant overall treatment effect we find could be the result of different elements of the program. The current set-up of the field experiment does not allow us to study the different elements separately. Future research with different treatment variations, e.g. in the team component and the entrepreneurial tasks in the program, are necessary to be able to disentangle the different effects.

### 3.4.2 Conclusion

Given the key role entrepreneurial activity has in fostering economic growth and innovation, the evaluation of measures that may stimulate successful entrepreneurship is

of the high interest to both academics and practitioners alike. Since entrepreneurship education programs are used worldwide, we thus believe that testing their effectiveness is an important first step. The evaluation studies that have been performed so far have only found modest effects at most as well as contradictory results. This seems to suggest that these programs are ineffective as a policy tool to promote entrepreneurial knowledge, skills or intentions.

However, until now the focus has been on entrepreneurship programs targeted at adolescents in secondary or higher education. The insignificant effects found there may well be due to the fact that entrepreneurial skills and knowledge are more easily developed earlier in life or because the returns to training programs later in life depend on investments in knowledge and skills made earlier. In fact, the model of skill formation introduced by Cunha and Heckman (2007) emphasizes such dynamic spill-over effects. In this model cognitive and non-cognitive skills are developed during different stages in life, where the skills learned during one period in life (e.g. at primary school) augment the benefits of investments in these skills in subsequent periods (e.g. at high school or university). Early investments in skills may thus be particularly effective in the long run.

In view of the potential importance of *early* educational investments, we evaluate the immediate (short term) effect of entrepreneurship education on the development of entrepreneurship knowledge and non-cognitive entrepreneurial skills of children aged 11 or 12. We also consider the program's impact on entrepreneurial intentions. By using a randomized field experiment we are able to obtain unbiased estimates. Our main finding indicates that the program has the intended effect; pupils in the treatment group show a significant increase in their non-cognitive entrepreneurial skills compared to those in the control group. Entrepreneurship knowledge is unaffected by the program though. The negative effects on entrepreneurial intentions must be taken with a pinch of salt, because measuring entrepreneurial intentions of children at the age of 11 or 12 is difficult. However, as pointed out by von Graevenitz et al. (2010), an overall decline in entrepreneurial intentions might actually be the preferred response to the program. If the program provides the children with a more realistic view of what it entails to be an entrepreneur, this could cause a positive sorting effect in that only those pupils with high entrepreneurial ability will choose an entrepreneurial career.

The program evaluated in this study takes five days and has a significant and quite substantial positive effect on the development of non-cognitive entrepreneurial skills. Remarkably, the program aimed at college students evaluated by Oosterbeek et al. (2010) is more involved in both time and costs and has no discernible effect on entrepreneurial skill development. Moreover, as mentioned above, the skills formation literature inspired by Cunha and Heckman (2007) strongly suggests that there are

important dynamic spill-over effects in the development of skills over time. It may therefore be likely that the effects of entrepreneurship programs in tertiary education will become larger among people who participated in these programs at a younger age. Additionally, the early development of non-cognitive skills may have a wider impact because they are known to have a positive effect on labor market outcomes in general. It thus appears that non-cognitive entrepreneurial skills are best developed already at an early age.

The positive results are novel and remarkable, even though they reflect only effects in the short run from one specific entrepreneurship education program. Obviously, our design does not allow the measurement of longer term effects of early entrepreneurship education because all children eventually participated in the program (justified on ethical grounds). Nevertheless, finding short term effects is a first step towards a better understanding of the effects of entrepreneurship education and the validity of dynamic spillover effects in the realm of entrepreneurship education.

A word of caution is required with respect to the exact mechanism that drives our results. The qualitative evidence discussed in Section 3.4.1 show that teamwork, in addition to (or instead of) the entrepreneurship element of the program, might be an important factor in the development of the non-cognitive skills. The current set-up of our field experiment unfortunately prevents us from disentangling the effect of these two mechanisms. Hence, we must leave it to future research to determine which of these elements has the greatest impact on the development of non-cognitive entrepreneurial skills. Another important drawback of our research design is that we do not measure the opportunity cost of the program. Even though the program only lasts five days, it would be interesting to know if, and to what extent, participation crowds out the learning of other types of knowledge and skills.

Few studies have so far employed methods that allow a similar causal interpretation. We only evaluate one specific early entrepreneurship program and Oosterbeek et al. (2010) evaluate only one specific program aimed at college students. It may well be the case that results for other programs are different (although both of these programs are the largest in their league worldwide). Hence, the results only suggest that early entrepreneurship education is more effective than later entrepreneurship education. The significant immediate (short term) impact on non-cognitive entrepreneurial skills for children established in this paper may be encouraging for (entrepreneurship) education policy. Our result also provides a relevant first step for future research to investigate whether the model of skill formation indeed holds for the development of entrepreneurial skills as well.

# Appendix A

Table A3.1: Composition of classes within the schools across years

| # schools | 2009-2010 or 2010-2011 | 2009-2010 *and* 2010-2011 |
|---|---|---|
| 1 class | 29 | 10 |
| 2 classes | 9 | 11 |
| $>$ 2 classes | 2 | 2 |

Table A3.2: Treatment effects Entrepreneurship knowledge and Intentions (detailed)

| | *Entrepreneurial intentions* | | |
|---|---|---|---|
| $\Delta$ | Future job: entrepreneur | Own Business | Entrepreneurship knowledge |
| Treatment effect ($\delta$) | -0.08 (0.05) | -0.22 (0.05) | 0.04 (0.06) |
| **Background (individual)** | | | |
| Female | -0.18 (0.04) | -0.12 (0.04) | 0.11 (0.04) |
| Age ($t = 0$) | -0.02 (0.02) | -0.02 (0.02) | -0.02 (0.02) |
| Parents both not dutch | -0.07 (0.04) | -0.03 (0.05) | -0.15 (0.05) |
| Mother entrepreneur | 0.11 (0.07) | 0.16 (0.06) | 0.06 (0.06) |
| Father entrepreneur | 0.16 (0.05) | 0.25 (0.05) | -0.05 (0.05) |
| Intention level at $t = 0$ | -1.27 (0.04) | -0.54 (0.02) | |
| Knowledge level at $t = 0$ | | | -0.79 (0.03) |
| *High school:* | | | |
| Pre-University | 0.20 (0.06) | 0.09 (0.05) | 0.76 (0.06) |
| Pre-Uni and senior general | 0.11 (0.06) | 0.13 (0.05) | 0.66 (0.06) |
| Senior general secondary | 0.11 (0.06) | 0.003 (0.07) | 0.42 (0.07) |
| Pre-vocational and senior general | 0.09 (0.06) | 0.10 (0.05) | 0.36 (0.07) |
| *(omitted category: Pre-vocational)* | | | |
| **Background (school)** | | | |
| Class size | -0.005 (0.02) | 0.03 (0.03) | 0.03 (0.03) |
| Avg. income per year (x €1000, −) | -0.02 (0.02) | 0.02 (0.02) | 0.008 (0.02) |
| Protestant | -0.02 (0.05) | -0.04 (0.05) | 0.09 (0.08) |
| Roman Catholic | -0.08 (0.05) | -0.04 (0.05) | 0.07 (0.08) |
| Religion other | 0.18 (0.10) | 0.07 (0.10) | 0.03 (0.13) |
| Year dummy (1= 2010/0=2011) | 0.02 (0.04) | -0.03 (0.04) | -0.03 (0.05) |
| _cons | 0.36 (0.06) | 0.15 (0.07) | -0.40 (0.09) |
| **Number of observations** | 2360 | 2354 | 2141 |

Robust standard errors in parentheses. Observations clustered at the class level.

Table A3.3: Results including team treatment dummies

| | Risk taking | Creativity | Need for Ach | Self-efficacy | Social orient. | Pro-activity | Persistence | Analyzing | Motivating | Ent. know. | Entrepreneur | Own Business |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Team composition** | | | | | | | | | | | | |
| JAT team | 0.101 | -0.097 | -0.043 | -0.049 | 0.113 | -0.065 | 0.011 | -0.004 | -0.034 | -0.004 | -0.017 | 0.053 |
| | (0.098) | (0.120) | (0.098) | (0.099) | (0.097) | (0.071) | (0.086) | (0.106) | (0.107) | (0.030) | (0.037) | (0.044) |
| Mixed spec. team | 0.021 | 0.081 | -0.029 | 0.035 | 0.128 | 0.110 | 0.034 | 0.028 | 0.026 | 0.004 | 0.017 | 0.002 |
| | (0.109) | (0.118) | (0.123) | (0.109) | (0.121) | (0.079) | (0.097) | (0.118) | (0.124) | (0.031) | (0.048) | (0.044) |
| Math spec. team | -0.012 | -0.117 | -0.143 | -0.031 | 0.074 | -0.056 | -0.141 | 0.012 | 0.101 | 0.017 | -0.001 | -0.073 |
| | (0.161) | (0.131) | (0.121) | (0.122) | (0.124) | (0.086) | (0.111) | (0.122) | (0.131) | (0.036) | (0.055) | (0.081) |
| Verbal spec. team | -0.045 | -0.032 | 0.109 | -0.126 | 0.117 | -0.086 | 0.005 | -0.031 | -0.148 | -0.001 | 0.002 | -0.086 |
| | (0.169) | (0.147) | (0.116) | (0.131) | (0.133) | (0.114) | (0.123) | (0.144) | (0.134) | (0.037) | (0.064) | (0.059) |
| *(ommitted category: left over teams)* | | | | | | | | | | | | |
| Constant | 2.785*** | 1.770* | 1.619** | 1.919** | 1.978** | 1.666* | 1.093 | 1.561* | 1.223 | 0.319 | 0.164 | 0.527 |
| | (0.553) | (0.826) | (0.592) | (0.697) | (0.601) | (0.655) | (0.632) | (0.646) | (0.623) | (0.162) | (0.257) | (0.349) |
| Background (ind.) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Background (school) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Number of obs.** | 986 | 982 | 986 | 986 | 986 | 985 | 986 | 986 | 985 | 978 | 1019 | 1018 |
| $R^2$ | 0.199 | 0.211 | 0.181 | 0.216 | 0.233 | 0.197 | 0.220 | 0.250 | 0.218 | 0.343 | 0.325 | 0.315 |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The tests in this table are only performed for the treatment group in 2010 because the control group is unaffected by the treatment variations and there was no other treatment variation in 2011. Standard errors in parentheses

## Table A3.4: Results including CSR treatment dummies

| | Risk taking | Creativity | Need for Ach | Self-efficacy | Social orient | Pro-activity | Persistence | Analyzing | Motivating | Ent. know. | Entrepreneur | Own Business |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Reward structure** | | | | | | | | | | | | |
| incentivized CSR | -0.109 | 0.002 | -0.169 | -0.073 | 0.151 | -0.017 | -0.123 | 0.107 | -0.043 | -0.010 | 0.008 | 0.065 |
| | (0.091) | (0.100) | (0.090) | (0.101) | (0.108) | (0.095) | (0.107) | (0.090) | (0.117) | (0.034) | (0.042) | (0.093) |
| non-incentivized CSR | 0.015 | 0.198* | 0.015 | 0.029 | 0.159 | -0.025 | 0.004 | 0.178 | -0.034 | 0.006 | -0.015 | -0.028 |
| | (0.074) | (0.092) | (0.086) | (0.096) | (0.092) | (0.077) | (0.091) | (0.112) | (0.101) | (0.033) | (0.039) | (0.070) |
| *(ommitted category: baseline)* | | | | | | | | | | | | |
| Constant | 2.822*** | 1.866* | 1.793** | 1.997** | 1.918** | 1.755** | 1.177 | 1.547* | 1.143 | 0.323 | 0.163 | 0.436 |
| | (0.557) | (0.857) | (0.592) | (0.705) | (0.651) | (0.641) | (0.646) | (0.675) | (0.624) | (0.162) | (0.246) | (0.356) |
| Background (ind.) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Background (school) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Number of obs.** | 988 | 984 | 988 | 988 | 988 | 987 | 988 | 988 | 987 | 979 | 1021 | 1020 |
| $R^2$ | 0.200 | 0.213 | 0.184 | 0.216 | 0.234 | 0.193 | 0.222 | 0.251 | 0.216 | 0.343 | 0.324 | 0.315 |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The tests in this table are only performed for the treatment group in 2010 because the control group is unaffected by the treatment variations and there was no other treatment variation in 2011. Standard errors in parentheses

Table A3.5: Treatment effects Non-cognitive entrepreneurial skills (detailed)

| $\Delta$ | Motivating | Analyzing | Pro-activity | Creativity | Self-efficacy | Need for Ach | Risk taking | Social orient | Persistence |
|---|---|---|---|---|---|---|---|---|---|
| Treatment effect ($\delta$) | 0.06 (0.05) | 0.14 (0.05) | 0.17 (0.05) | 0.10 (0.05) | 0.16 (0.04) | 0.15 (0.05) | 0.11 (0.04) | 0.05 (0.05) | 0.10 (0.05) |
| **Background (individual)** | | | | | | | | | |
| Female | 0.06 (0.04) | -0.08 (0.03) | 0.004 (0.04) | -0.06 (0.04) | -0.07 (0.04) | -0.09 (0.04) | -0.11 (0.03) | 0.11 (0.04) | 0.07 (0.03) |
| Age ($t = 0$) | 0.02 (0.02) | 0.02 (0.02) | 0.008 (0.02) | 0.01 (0.02) | 0.02 (0.02) | -0.01 (0.02) | -0.003 (0.02) | -0.01 (0.02) | 0.03 (0.02) |
| Parents both not dutch | 0.05 (0.05) | 0.02 (0.06) | -0.02 (0.05) | -0.004 (0.05) | 0.05 (0.05) | 0.04 (0.05) | -0.05 (0.05) | -0.05 (0.06) | 0.05 (0.05) |
| Mother entrepreneur | -0.07 (0.09) | 0.02 (0.07) | 0.08 (0.06) | 0.10 (0.07) | -0.05 (0.07) | 0.04 (0.06) | 0.05 (0.07) | 0.03 (0.07) | -0.03 (0.07) |
| Father entrepreneur | 0.009 (0.06) | 0.02 (0.05) | 0.04 (0.04) | 0.03 (0.05) | -0.009 (0.05) | 0.13 (0.05) | 0.02 (0.05) | 0.008 (0.05) | -0.03 (0.04) |
| Skill level at $t = 0$ | -0.47 (0.02) | -0.51 (0.02) | -0.46 (0.02) | -0.47 (0.02) | -0.46 (0.02) | -0.46 (0.02) | -0.44 (0.02) | -0.46 (0.02) | -0.46 (0.02) |
| *High school:* | | | | | | | | | |
| Pre-University | 0.22 (0.06) | 0.48 (0.07) | 0.33 (0.06) | 0.33 (0.06) | 0.28 (0.06) | 0.26 (0.06) | 0.04 (0.06) | 0.09 (0.05) | 0.27 (0.06) |
| Pre-Uni and senior general | 0.22 (0.06) | 0.36 (0.06) | 0.21 (0.06) | 0.30 (0.06) | 0.15 (0.06) | 0.25 (0.06) | 0.14 (0.06) | 0.13 (0.06) | 0.22 (0.06) |
| Senior general secondary | 0.20 (0.06) | 0.13 (0.07) | 0.19 (0.07) | 0.20 (0.07) | 0.15 (0.07) | 0.12 (0.06) | 0.02 (0.06) | 0.10 (0.06) | 0.10 (0.07) |
| Pre-vocational and senior general | 0.22 (0.06) | 0.04 (0.07) | 0.04 (0.07) | 0.08 (0.07) | 0.05 (0.07) | 0.08 (0.06) | 0.11 (0.06) | 0.01 (0.06) | 0.17 (0.06) |
| *(omitted category: Pre-vocational)* | | | | | | | | | |
| **Background (school)** | | | | | | | | | |
| Class size | -0.007 (0.02) | 0.02 (0.02) | -0.03 (0.02) | -0.01 (0.02) | -0.04 (0.03) | -0.004 (0.03) | -0.02 (0.02) | 0.002 (0.03) | 0.008 (0.03) |
| Avg. income per year (x €1000, −) | 0.02 (0.02) | -0.003 (0.02) | 0.01 (0.02) | -0.02 (0.02) | 0.01 (0.02) | 0.01 (0.03) | 0.02 (0.01) | 0.03 (0.03) | 0.02 (0.02) |
| Protestant | 0.02 (0.05) | -0.02 (0.05) | 0.08 (0.05) | 0.004 (0.05) | -0.04 (0.05) | -0.03 (0.05) | -0.04 (0.05) | -0.06 (0.06) | -0.06 (0.05) |
| Roman Catholic | 0.09 (0.05) | -0.02 (0.05) | 0.03 (0.05) | -0.008 (0.06) | 0.01 (0.06) | 0.02 (0.06) | -0.03 (0.04) | 0.005 (0.06) | -0.07 (0.05) |
| Religion other | 0.17 (0.07) | 0.07 (0.13) | 0.25 (0.10) | 0.11 (0.10) | 0.27 (0.08) | 0.04 (0.08) | 0.12 (0.08) | -0.07 (0.10) | 0.08 (0.08) |
| Year dummy (1= 2010/0=2011) | 0.0004 (0.05) | -0.002 (0.05) | 0.04 (0.04) | -0.01 (0.04) | 0.08 (0.05) | -0.02 (0.05) | 0.0006 (0.04) | -0.03 (0.05) | 0.006 (0.04) |
| _cons | -0.28 (0.07) | -0.26 (0.08) | -0.35 (0.08) | -0.23 (0.06) | -0.25 (0.08) | -0.22 (0.07) | -0.06 (0.06) | -0.10 (0.08) | -0.23 (0.07) |
| **Number of observations** | 2302 | 2304 | 2303 | 2297 | 2304 | 2304 | 2304 | 2304 | 2304 |

Robust standard errors in parentheses. Observations clustered at the class level.

Table A3.6: Treatment effects Non-cognitive entrepreneurial skills (clustered at school level)

| $\Delta$ | Motivating | Analyzing | Pro-activity | Creativity | Self-efficacy | Need for Ach | Risk taking | Social orient | Persistence |
|---|---|---|---|---|---|---|---|---|---|
| Treatment effect ($\delta$) | 0.07 (0.05) | 0.14 (0.05) | 0.17 (0.05) | 0.11 (0.05) | 0.16 (0.04) | 0.16 (0.06) | 0.12 (0.05) | 0.05 (0.05) | 0.11 (0.05) |
| **Background (individual)** | | | | | | | | | |
| Female | 0.07 (0.04) | -0.08 (0.03) | 0.004 (0.03) | -0.06 (0.04) | -0.07 (0.04) | -0.10 (0.05) | -0.12 (0.04) | 0.11 (0.04) | 0.08 (0.04) |
| Age ($t = 0$) | 0.03 (0.03) | 0.03 (0.04) | 0.01 (0.03) | 0.02 (0.04) | 0.02 (0.04) | -0.02 (0.03) | -0.005 (0.03) | -0.02 (0.03) | 0.06 (0.03) |
| Parents both not dutch | 0.05 (0.06) | 0.02 (0.05) | -0.02 (0.05) | -0.005 (0.07) | 0.05 (0.05) | 0.04 (0.06) | -0.06 (0.05) | -0.05 (0.05) | 0.06 (0.05) |
| Mother entrepreneur | -0.08 (0.11) | 0.02 (0.06) | 0.08 (0.06) | 0.11 (0.07) | -0.05 (0.07) | 0.05 (0.07) | 0.05 (0.07) | 0.03 (0.07) | -0.03 (0.07) |
| Father entrepreneur | 0.01 (0.06) | 0.02 (0.05) | 0.04 (0.05) | 0.03 (0.05) | -0.009 (0.04) | 0.14 (0.04) | 0.02 (0.06) | 0.008 (0.05) | -0.03 (0.04) |
| Competency at $t = 0$ | -0.45 (0.02) | -0.45 (0.02) | -0.43 (0.02) | -0.42 (0.02) | -0.43 (0.02) | -0.37 (0.02) | -0.38 (0.02) | -0.41 (0.02) | -0.43 (0.02) |
| *High school:* | | | | | | | | | |
| Pre-University | 0.26 (0.08) | 0.47 (0.07) | 0.33 (0.06) | 0.39 (0.06) | 0.28 (0.06) | 0.27 (0.06) | 0.05 (0.07) | 0.09 (0.06) | 0.29 (0.06) |
| Pre-Uni and senior general | 0.26 (0.07) | 0.35 (0.06) | 0.21 (0.06) | 0.34 (0.07) | 0.15 (0.06) | 0.26 (0.06) | 0.16 (0.07) | 0.13 (0.05) | 0.24 (0.06) |
| Senior general secondary | 0.23 (0.08) | 0.13 (0.07) | 0.19 (0.06) | 0.23 (0.08) | 0.15 (0.08) | 0.13 (0.06) | 0.02 (0.06) | 0.10 (0.05) | 0.10 (0.06) |
| Pre-vocational and senior general | 0.25 (0.07) | 0.04 (0.08) | 0.04 (0.07) | 0.10 (0.09) | 0.05 (0.06) | 0.08 (0.07) | 0.12 (0.07) | 0.01 (0.06) | 0.18 (0.07) |
| **Background (school)** | | | | | | | | | |
| Class size | -0.002 (0.01) | 0.004 (0.01) | -0.007 (0.00) | -0.003 (0.01) | -0.01 (0.01) | -0.0008 (0.01) | -0.005 (0.01) | 0.0005 (0.01) | 0.002 (0.01) |
| Avg. income per year (x €1000, −) | 0.006 (0.01) | -0.0008 (0.00) | 0.003 (0.00) | -0.005 (0.01) | 0.003 (0.01) | 0.003 (0.01) | 0.005 (0.00) | 0.006 (0.01) | 0.006 (0.01) |
| Protestant | 0.03 (0.06) | -0.02 (0.05) | 0.08 (0.05) | 0.005 (0.05) | -0.04 (0.06) | -0.03 (0.06) | -0.04 (0.05) | -0.06 (0.05) | -0.07 (0.05) |
| Roman Catholic | 0.10 (0.06) | -0.02 (0.05) | 0.03 (0.05) | -0.009 (0.06) | 0.01 (0.06) | 0.02 (0.07) | -0.03 (0.05) | 0.005 (0.06) | -0.08 (0.05) |
| Religion other | 0.20 (0.07) | 0.07 (0.06) | 0.25 (0.05) | 0.12 (0.07) | 0.27 (0.05) | 0.05 (0.08) | 0.14 (0.08) | -0.07 (0.07) | 0.08 (0.05) |
| Year dummy (1= 2010/0=2011) | 0.0005 (0.05) | -0.002 (0.04) | 0.04 (0.04) | -0.01 (0.05) | 0.08 (0.04) | -0.02 (0.05) | 0.0007 (0.04) | -0.03 (0.05) | 0.006 (0.05) |
| _cons | 1.56(0.42) | 1.38 (0.46) | 1.66 (0.45) | 1.72 (0.48) | 1.60 (0.49) | 1.85 (0.44) | 1.85 (0.42) | 2.19 (0.41) | 1.02 (0.42) |
| **Number of observations** | 2302 | 2304 | 2303 | 2297 | 2304 | 2304 | 2304 | 2304 | 2304 |

Robust standard errors in parentheses. Observations clustered at the school level.

Table A3.7: Treatment effects Entrepreneurship knowledge and intentions (clustered at school level)

| | *Entrepreneurial intentions* | | |
|---|---|---|---|
| $\Delta$ | Future job: entrepreneur | Own Business | Entrepreneurship knowledge |
| Treatment effect ($\delta$) | -0.03 (0.02) | -0.14 (0.03) | 0.01 (0.01) |
| **Background (individual)** | | | |
| Female | -0.08 (0.02) | -0.08 (0.02) | 0.02 (0.01) |
| Age ($t = 0$) | -0.02 (0.01) | -0.02 (0.02) | -0.006 (0.01) |
| Parents both not dutch | -0.03 (0.01) | -0.02 (0.03) | -0.03 (0.01) |
| Mother entrepreneur | 0.05 (0.04) | 0.10 (0.04) | 0.01 (0.01) |
| Father entrepreneur | 0.08 (0.02) | 0.16 (0.03) | -0.01 (0.01) |
| Intention level at $t = 0$ | -0.58 (0.02) | -0.56 (0.02) | |
| Knowledge level at $t = 0$ | | | -0.72 (0.03) |
| *High school:* | | | |
| Pre-University | 0.09 (0.03) | 0.06 (0.04) | 0.16 (0.01) |
| Pre-Uni and senior general | 0.05 (0.03) | 0.08 (0.04) | 0.14 (0.01) |
| Senior general secondary | 0.05 (0.03) | 0.002 (0.04) | 0.09 (0.01) |
| Pre-vocational and senior general | 0.04 (0.03) | 0.06 (0.03) | 0.09 (0.02) |
| **Background (school)** | | | |
| Class size | -0.0006 (0.00) | 0.004 (0.00) | 0.001 (0.00) |
| Avg. income per year (x €$1000, -$) | -0.002 (0.00) | 0.003 (0.00) | 0.001 (0.00) |
| Protestant | -0.007 (0.02) | -0.03 (0.03) | 0.02 (0.02) |
| Roman Catholic | -0.03 (0.02) | -0.03 (0.03) | 0.01 (0.02) |
| Religion other | 0.08 (0.04) | 0.05 (0.05) | 0.01 (0.02) |
| Year dummy (1= 2010/0=2011) | 0.01 (0.02) | -0.02 (0.03) | -0.006 (0.01) |
| _cons | 0.43 (0.17) | 0.74 (0.25) | 0.50 (0.11) |
| **Number of observations** | 2360 | 2354 | 2190 |

Robust standard errors in parentheses. Observations clustered at the school level.

Table A3.8: Treatment effects using standardized measures (class level analysis)

| | **DID no controls** | | **DID with controls** | |
|---|---|---|---|---|
| **Outcome variables** | $\delta$ | | $\delta$ | |
| *Non-cognitive entrepreneurial skills* | | | | |
| Risk Taking | 0.064 | (0.05) | 0.073 | (0.05) |
| Creativity | 0.046 | (0.05) | 0.046 | (0.05) |
| Need for Achievement | 0.144*** | (0.05) | 0.130** | (0.05) |
| Self-Efficacy | 0.117** | (0.05) | 0.126** | (0.05) |
| Social Orientation | 0.036 | (0.06) | 0.001 | (0.06) |
| Pro-activity | 0.123** | (0.05) | 0.170*** | (0.05) |
| Persistence | 0.082* | (0.05) | 0.090* | (0.05) |
| Analyzing | 0.125** | (0.05) | 0.135** | (0.06) |
| Motivating | 0.034 | (0.05) | 0.026 | (0.05) |
| *Entrepreneurship knowledge* | 0.073 | (0.07) | 0.058 | (0.08) |
| *Entrepreneurial intentions* | | | | |
| Future job: entrepreneur (0/1) | -0.081 | (0.05) | -0.058 | (0.05) |
| Own Business (0-2) | -0.247*** | (0.05) | -0.252*** | (0.06) |
| **Number of observations** | 104 | | 104 | |

Note: The estimates in each cell come from separate regressions, robust standard errors are in parentheses. All regressions control for the baseline level of the outcome variable. DID with controls includes individual characteristics: age, gender, future high school track, nationality parents, parents entrepreneurial status; school/neighborhood characteristics: class size, school denomination, avg. income per year and a year dummy for 2010/2011. */**/*** indicates significance at the 10%/5%/1%-level.

# Appendix B – Translation questionnaire

| Part 1 (background characterics) | Pre-test | Post-test |
|---|---|---|
| First name | x | x |
| Last name | x | x |
| Date of birth | x | x |
| Boy or girl | x | x |
| Zipcode | x | x |
| How many older siblings do you have? | x | |
| How many younger siblings do you have? | x | |
| Do your parents live together? | x | |
| Are you a member of the public library? | x | |
| Do you have a subscription to a newspaper at home? | x | |
| Student number (given to you by the teacher) | x | x |
| Do you sometimes go to a museum with your parents or other family members? | x | |
| Do you sometimes go to the theater? | x | |
| Do you sometimes get a book as a present? | x | |
| Below is a list of occupations. Please choose at most three that you would like to become later. (22 categories) | x | x |
| Which high school track will you probably go to after primary school? | x | |
| Do you think that you would like to start your own company one day? | x | x |
| Do you think that you would like to become a manager in an existing company one day? | x | x |
| Do you think that you would like to become an employee (work for a manager) one day? | x | x |
| Do you think you would like to go to university after high school? | x | x |
| What is your favorite class in school? | x | x |
| What is the country of origin of your mother and father? | x | |
| Does your mother have a job? | x | |
| If so, what is her occupation? | x | |
| What type of tasks are part of your mother's work? (7 options) | x | |
| What is your mother's education level? (five categories) | x | |
| Does your father have a job? | x | |
| If so, what is his occupation? | x | |
| What type of tasks are part of your father's work? (7 options) | x | |
| What is your father's education level? (five categories) | x | |
| If you are allowed to choose between a verbal task and a mathematical task, which one would you prefer? | x | |
| Are better at mathematical or at verbal tasks? | x | |
| What are your actual hobbies? (24 categories) | x | |
| What hobbies would you have if everything was allowed? (24 categories) | x | |
| Do you get pocket money? | x | |
| If so, how much do you get? | x | |
| Do you get this pocket money per month or per week? | x | |
| Are there things that you have to buy with your own pocket money? (12 categories) | x | |
| Imagine that you have done a chore for someone in order to earn money, like mowing the lawn or painting a wall. Which of the following pay out options would you choose? (4 categories) | x | x |
| Below is a list of occupations. Please grade each occupation (on a scale from 0-10) according to how much you value this job or how important it | x | x |

| | | |
|---|---|---|
| is to you. (22 categories) | | |
| Is your father an entrepreneur (works in his own company)? | | x |
| Is your mother an entrepreneur (works in her own company)? | | x |
| | | |
| Part 2 (non-cognitive skills) | Pre-test | Post-test |
| Below is a list of statements. Please indicate for each statement (on a scale of 1-7) how much this statement applies to you or suits you.<br>1 = Does not apply to me at all<br>7 = Suits me perfectly<br>So if a statement applies to you more or suits you better, you should choose a higher grade. | | |
| I can list and weigh advantages and disadvantages well | x | x |
| I dislike unfinished work | x | x |
| I don't mind taking risks | x | x |
| If there is a problem, I immediately try to solve it | x | x |
| I am able to understand difficult things | x | x |
| I like talking to children that I haven't met before | x | x |
| I want to show what I am capable of | x | x |
| I have a vivid imagination | x | x |
| I think I am good at solving problems | x | x |
| I know what a good company is | x | x |
| If I start something, I go on until it is done | x | x |
| I like to put my ideas into actions | x | x |
| I am able to solve a difficult puzzle quickly | x | x |
| I want to perform better than others | x | x |
| I am often one of the first to have a good idea | x | x |
| No matter what happens, I am able to handle it | x | x |
| I often come up with original solutions | x | x |
| I want to achieve things that others cannot achieve | x | x |
| I often come up with new ideas | x | x |
| I get along well with other children | x | x |
| I am able to see if people are good or bad at their job | x | x |
| I like to take chances | x | x |
| I notice if other children are enjoying themselves | x | x |
| I am able to get other children to participate | x | x |
| If I start something new, I know I will succeed | x | x |
| I can encourage other children to do their best | x | x |
| I know what an entrepreneur does or what entrepreneurship means | x | x |
| I dare to take risks, even if I could lose | x | x |
| If there is something that I cannot do, I keep practicing until I can do it | x | x |
| I am able to make others enthusiastic | x | x |
| | | |
| Part 3 (Entrepreneurship knowledge) | Pre-test | Post-test |
| If a company earns more money by selling products or services than it spends, it will...<br>a) be registered at the stock market<br>b) make a profit<br>c) make a loss<br>d) have debt | x | |
| If a company earns less money by selling products or services than it spends, it will...<br>e) be registered at the stock market<br>f) make a profit<br>g) make a loss | | x |

| | | |
|---|---|---|
| h) have debt | | |
| When a company sells shares to a venture capitalist, it exchanges part of its ownership for...<br>a) employees<br>b) products or services<br>c) money<br>d) benefits for employees | x | |
| If people work for a company they receive...<br>a) a loan<br>b) a salary<br>c) stocks<br>d) debt | x | x |
| What do banks get if they lend money to a company?<br>a) shares in the company<br>b) a promise that the loan will be redeemed with interest<br>c) part of the profits of the company<br>d) discount on the products of the company | x | |
| To set the price for a product you have to take the following into account<br>a) how much it costs to make a product<br>b) how many products can be made in a certain amount of time<br>c) the price that competitors ask for their products<br>d) all of the above answers are correct | x | x |
| Advertisements, like TV commercials, posters, logo's and a slogan<br>a) make sure that the company will be successful<br>b) always feature a famous person<br>c) try to convince people to buy the products of the company<br>d) all of the above answers are correct | x | x |
| It is important for a company to<br>a) have a good customer service<br>b) offer good products and services<br>c) be honest in advertisements<br>d) all of the above answers are correct | x | x |
| A company earn revenues by selling<br>a) employees<br>b) products or services<br>c) money<br>d) benefits of the employees | | x |
| What does a venture capitalist get if he invests money in a company?<br>e) shares in the company<br>f) a promise that the loan will be redeemed with interest<br>g) part of the profits of the company<br>h) discount on the products of the company | | x |
| | | |
| Part 1b (Questions about BizWorld) | Pre-test | Post-test |
| How did you like BizWorld? (5 categories) | | x |
| What grade would you give BizWorld? (0-10) | | x |
| Which BizWorld day did you like best? | | x |
| Which part of BizWorld did you like best? (8 categories) | | x |
| What was your role within the team? (6 categories) | | x |
| What grade would you give to your own performance? (0-10) | | x |
| What grade would you give to your own effort? (0-10) | | x |
| What grade would you give to your own motivation? (0-10) | | x |
| How well did you perform during BizWorld? (5 categories) | | x |
| What was the most important reason for you to perform well? (7 categories) | | x |

| | | |
|---|---|---|
| Did your team register a patent? | | x |
| How much money did your team borrow as a company? (5 categories) | | x |
| How many shares did you sell? (0-10) | | x |
| What grade would you give to your team's performance? (0-10) | | x |
| What grade would you give to your team's effort? (0-10) | | x |
| What grade would you give to your team's motivation? (0-10) | | x |
| Team name | | x |
| How many children were there on your team (including yourself)? | | x |
| How well did you work together as a team? (5 categories) | | x |
| How many conflicts were there in your team during the entire BizWorld program? | | x |
| Were you absent on one of the BizWorld days? | | x |
| What was the name of the team that won in your class? | | x |
| Do you think the winning team deserved to win? (5 categories) | | x |
| What do you think about the option to produce sustainable products? (5 categories) | | x |
| If your team chose the sustainability option, why did you do that? (6 categories) | | x |
| What was the mission of your team? | | x |
| What determines the company value of your team? | | x |
| Did you choose to apply for the sustainability trademark? | | x |
| Do you have any comments on how to improve the BizWorld program? | | x |
| What do you think about the teaching of the trainer in your class? (8 categories) | | x |
| What grade would you give him/her? (0-10) | | x |

# Chapter 4

# Balanced skills and team performance

*This chapter is based on Huber, L.R., R. Sloof and M. van Praag (2014), Jacks-of-all-trades? The effect of balanced skills on team performance, IZA Discussion Paper No. 8237.*

## 4.1 Introduction

Nowadays, teamwork is an omnipresent phenomenon within firms of all types and sizes (Hamilton et al., 2003). Large firms increasingly rely on the work and decisions made by (self-managed) teams (Lazear and Shaw, 2007). Moreover, a substantial and growing share of businesses are started up and run by entrepreneurial teams instead of solo entrepreneurs (Klotz et al., 2014; Parker, 2009). Over the past decades the tasks within (large) established organizations have become increasingly complex due to new technologies and rapidly changing environments. Complex environments require the combination of different types of knowledge in order to be successful (Hoever et al., 2012; Dahlin et al., 2005). Similarly, the creation of a new venture is a complex task in and of itself; to be able to bring people, ideas and physical resources together, entrepreneurs must have knowledge, at least at a basic level, of a large number of business areas (Lazear, 2005, 2004). Thus, in order to deal with the various tasks and challenges they are confronted with, entrepreneurs require a balanced set of skills, i.e., should be *Jacks-of-All-Trades* (JAT). The combination of an increase in task complexity and the observed increase in teamwork raises the question of effective team composition. In this chapter, we empirically explore this question focusing on the *skill* composition of successful teams.

The aim of this chapter is to answer two specific research questions: (1) Do teams

consisting of a combination of individual generalists (JATs) perform better than teams consisting of (one type of) individual specialists? and (2) Can a lack of skill balance at the individual level be compensated by combining different types of specialists within one team? These two questions are partly motivated by the findings presented by Rulke and Galaskiewicz (2000). They compare the performance of teams of generalists to the performance of teams of specialists using data on 34 teams of MBA students. They find that the teams of generalists outperform the specialist teams. Furthermore, they show that the difference in performance becomes smaller as the specialist teams become more decentralized, i.e., when knowledge sharing among team members becomes easier. However, the authors mention several drawbacks of their design that we will address in this chapter. First, one drawback of the study of Rulke and Galaskiewicz (2000) is that the definition of the (type of) specialists is based on self-reported measures. Second, they do not have objective measures of individual ability of the students, which could be a confounding factor when studying team performance. Finally, the composition of the teams in their experiment is endogenous, i.e. chosen by the (executive) manager of the team.

In their paper Rulke and Galaskiewicz (2000) do not distinguish between different types of specialist teams, i.e., homogeneous and heterogeneous specialist teams. We add to their study by distinguishing between two *types* of specialists. Based on these types we explicitly compose teams consisting of a combination of these types of specialists and compare their performance to the generalist teams. This enables us to answer our second research question about what is the most effective way to balance skills within a team. This question is also related to one of the predictions made by Lazear (2005). He argues that as production processes become more complex, the supply of suitable entrepreneurs, i.e. individuals who have a balanced set of all the relevant skills, will decrease. Teaming up with others in order to overcome a lack of skill balance at the individual level might be one way to overcome this problem.[1] Possibly the balanced set of skills needed to be successful as an entrepreneur does not have to be endowed within one person, but can be spread out over several members of an entrepreneurial team. In this setting not all team members need to have all the skills, as long as all the required skills are present within the team. Indeed, there is evidence that more and more start-ups are done by teams especially in settings where it is difficult to

[1] At the individual level, Lazear's Jack-of-all-trades theory has been tested in terms of entrepreneurship entry and performance. Positive results on entry have been found for example by Lazear (2005), Wagner (2006) and Lechmann and Schnabel (2011), while Silva (2007) shows that innate ability is more important than acquiring a balanced set of skills in the probability of becoming an entrepreneur. The relationship between balanced skills and performance is less well studied and the results are mixed. Hartog et al. (2010) find a positive effect of diversity in innate skills on entrepreneurial performance, whereas Åstebro and Thompson (2011) find that having a balanced skill set is driven by a taste for variety and is thereby negatively related to earnings.

have all the required skills endowed within one person, such as the high-tech industry (Lechler, 2001). Moreover, Lazear and Shaw (2007) argue that also within large firms, especially in environments of innovation and rapid changes, it becomes harder to find generalists that have a sufficient level of all relevant skills. Observational data confirms that within established firms the share of teamwork has increased substantially over the past decades (Lazear and Shaw, 2007, Table 1). Thus, also within this setting a potential solution could be to form a management team consisting of individual experts, instead of hiring one expert manager.

We conduct a field experiment to analyze the role and substitutability of balanced skills for teams within a dynamic (entrepreneurial) environment. We study teams of children who set up a toy business in friendship bracelets in an entrepreneurship education program ("BizWorld") in the last grade of primary school in the Netherlands. Within this setting the task that the teams have to perform has several dimensions that each affect team performance (e.g., the quality and the presentation of the business plan, production, sales, financial statements, etc). Two basic skills, verbal ability and mathematical ability, are (most) important at this age and thus are also deemed to be relevant for the tasks that have to be performed within the entrepreneurship education program.[2] Based on uniform and precisely measured scores on these skills for all 1,131 individuals, we exogenously compose 179 teams separated into four different team types: JAT-teams, (homogeneous) math and verbal specialist teams and mixed specialist teams. JAT-teams consist only of generalists, i.e. children with balanced skills at the individual level. Additionally, within these teams the average score on math is comparable (or equal to) the average score on verbal ability. Math- and verbal-specialist teams consist of children who are either math- or verbal-specialists. These are children that score higher on one of the two subjects compared to the other. The mixed-specialist teams consist of a combination of math- and verbal-specialists. However, just as the JAT-teams, within these teams average math ability (at the team level) is comparable to average verbal ability. We compare the performance of the different team types to determine to what extent the measured skill sets affect team performance.

The results show that JAT-teams perform significantly better than verbal-specialist teams. These results suggest that balanced skills are indeed beneficial for the performance of teams in this (complex) environment. Surprisingly however, the performance of math-specialist teams is not significantly different compared to the JAT-teams. To answer the second research question we compare JAT-teams and mixed-specialist

[2]In the commonly used Stanford-Binet test of abstract intelligence, verbal and quantitative reasoning (i.e. language and numerical ability) are two specific factors determining intelligence. Moreover, PISA (the OECD Program for International Student Assessment) distinguishes reading and mathematics as the two key subjects used for the worldwide evaluation of education systems (Source: www.oecd.org/pisa).

teams. Our results show that mixed-specialist teams perform significantly worse than JAT-teams. This indicates that it is hard to substitute individual balanced skills by combining different specialists within one team. Apparently the ability to combine resources effectively is not something that comes across when people combine their specialized skills within teams.

Obviously, conducting a field experiment in the BizWorld entrepreneurship education program provides a trade-off between internal and external validity. However, this particular setting and the controlled experimental design have various benefits which allow us to study an interesting causal effect. First and foremost, there is no self-selection bias. When studying real life (entrepreneurial) teams the causal effect of team composition on team performance is difficult to determine, among others due to self-selection into and out of teams. Moreover, the possibility of running a field experiment in a setting where actual employees or entrepreneurs are forced to work on a project or to start their business with a group of people randomly assigned to them is virtually impossible. Furthermore, all children in our sample have approximately the same age and exactly the same level and amount of education. This is important for two reasons. First, because the accurate measurement of the sheer effect of cognitive skills is problematic when education levels and ages differ (Heckman et al., 2006).[3] Second, the balancedness in skills for children at this age is exogenous, i.e. based on endowed skills instead of a selection of people who decided to invest in a diverse set of skills (Silva, 2007). Another benefit of this field experiment is that we are able to create a relevant and reliable measure of the balanced skill set. In the Netherlands mathematical and verbal ability are the two most important cognitive skills for the learning outcomes of children aged 11-12.[4] In the experiment these skills are measured using a uniform and valid scale shortly before the children enter the program. Obviously, when studying (venture) teams in actual practice, each business environment would require different sets of skills and the set of relevant skills that must be balanced would not be so easy to define or measure in a uniform way. Finally, in the BizWorld program, the relationship and interaction between team members is a crucial component of team performance. All in all, the education program provides us with a suitable real effort experiment.

Our paper contributes to the existing literature in several ways. First, there is little evidence on the effect of balanced skills at the team level on (team) performance. A few studies have been conducted to investigate balanced skills within new venture teams.

[3]Age affects measured ability, whereas there is reverse causality in the relationship between education and measured ability, i.e. schooling affects test scores and measured test scores predict schooling (Hansen et al., 2004, p.40).

[4]The Dutch Ministry of Education has defined verbal and mathematical skills as the two core skills for primary education in the Netherlands (http://www.onderwijsinspectie.nl/onderwerpen/Onderwijsinhoud/Basisvaardigheden)

For example, Colombo and Grilli (2005) show that there are possible gains from the combination of economic-managerial and scientific-technical education among members within a start-up team. Furthermore, Cantner et al. (2010) study two different types of functional heterogeneity in new venture teams: knowledge scope and knowledge disparity. The results show that a broad knowledge stock is important for new venture growth but has no influence on survival. Similarity or functional overlap is positively associated with setting up and maintaining a new venture but has no influence on venture growth. However, in these studies the composition of the teams is endogenous, i.e. chosen by the members of the team.

Hence, a second contribution of our paper is that the use of a field experiment with randomized assignment to teams enables us to establish a causal effect. Several other studies have studied the skill composition of teams in ways that permit causal inferences. For example, Kahane et al. (2013), Franck and Nüesch (2010) and Gould and Winter (2009) have studied the effect of skill dispersion on team outcomes using various professional sports settings. Moreover, Leonard et al. (2004) and Hamilton et al. (2003) have studied the effective composition in terms of skills relevant to production processes in retail and a garment factory, respectively. However, all these studies pertain to settings with simple (production) tasks instead of complex problem solving and decision making tasks, which is the focus of this chapter.

To the best of our knowledge, the only other paper that investigates the causal effect of skills diversity on team performance within a complex environment is the study by Hoogendoorn et al. (2012). Their paper differs from this paper in terms of skill dimension. Hoogendoorn et al. (2012) look at diversity in one dimension, i.e. ability, and estimate the effect of dispersed levels of cognitive ability (or IQ) on team performance. The composition of the teams in our sample and the definition of skill diversity are based on two dimensions, i.e., mathematical and verbal ability. Another contribution of our paper is that we explicitly study the intra-team substitutability of useful combinations of skills.

The remainder of this chapter is structured as follows. The research design is described in Section 4.2. Section 4.3 reports the empirical findings and in Section 4.4 we summarize and conclude.

## 4.2 Data and methodology

### 4.2.1 Design of the field experiment

The objective of this field experiment is to assess the effect of balanced skills for (entrepreneurial) teams. In the experiment teams of children perform tasks within the set-

ting of an entrepreneurship education program called BizWorld, which requires school children to set up a business in a team. The program also requires them to divide tasks by assigning distinct professional positions to each of the team members. Thus, this program provides us with an environment in which genuine teamwork is relevant and the tasks that have to be performed are complex and multidimensional.

According to the psychological literature, verbal and mathematical skills are the two core skills for children between the age of 6 and 15 years.[5] Arguably, all other skills developed later in life, are derived from these two core skills. For instance, analytical and technical ability are closely linked to mathematical ability, whereas communication and language skills can be associated with general verbal ability. The measurement of verbal and mathematical skills forms the core of a countrywide uniform exam that the children take just before the program starts. Hence, for both skills objective measures are available from this standardized test. The scores from these tests provide us with objective and comparable measures for (almost) all the children in the sample. At the stage of primary school where children have not yet been selected into various school types and levels, the variation in the scores of these skills is still maximal. Hence, based on an objectively measured and well defined skill set we are able to create a relevant and reliable measure of skill balance.

Moreover, the BizWorld program sets a clear and measurable objective for the participating teams, i.e., acquiring the highest ranking within your class. The exact way in which the ranking of the teams is determined is calculated for all teams at the same stage (after having completed the program) and will be described in more detail in Section 4.2.3. Finally, for the sake of the experiment we were allowed to compose teams based on the scores of the two skill types.

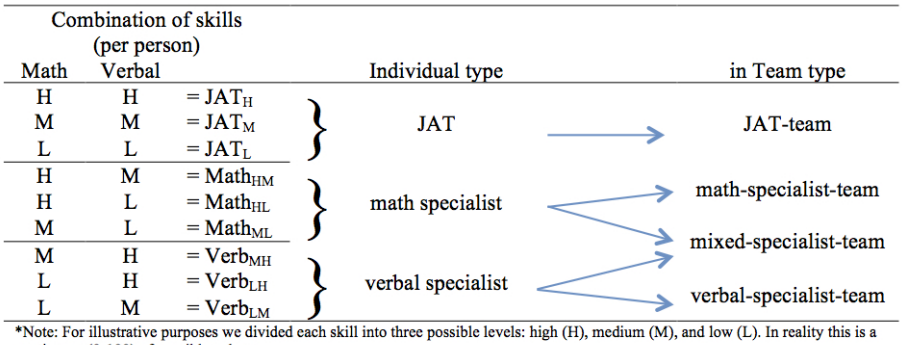
We compose four distinct team types using a two-step procedure. First, the children are classified as being one of three individual types based on their abilities. A child is classified as either a JAT, a math specialist or a verbal specialist (see Figure 4.1). We define JATs as those children with equal mathematical and verbal ability. We thereby closely follow Lazear's definition that JATs do not perform significantly better in one of the two tasks but they are equally good (or bad) in both (Lazear, 2005, p.656). Specialists are defined as pupils with a higher score in either math or verbal relative to the other subject.

For illustrative purposes, let us assume that for each skill there are only three possible levels: high (H), medium (M) or low (L), see Figure 4.1.[6] This would yield

[5] In the commonly used Stanford-Binet test of abstract intelligence, verbal and quantitative reasoning (i.e. language and numerical ability) are two specific factors determining intelligence. Moreover, PISA (the OECD Program for International Student Assessment) distinguishes reading and mathematics as the two key subjects used for the worldwide evaluation of education systems (Source: www.oecd.org/pisa).

[6] In reality we measure a continuum (0-100) of possible values. The measurement of the skills in

**Figure 4.1.** Team composition: from individuals to teams




*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.

three types of JATs: high, medium and low ability and, similarly, three types of math- and verbal specialists. In a second step we use the individual types to compose the four different team types: JAT-teams, mixed-specialist-teams, math-specialist-teams and verbal-specialist-teams. As can be seen in Figure 4.1, the JAT-teams consist only of JATs and therefore have balanced skills both at the individual and at the team level, i.e. within these teams the average scores on math- and verbal ability are comparable. The mixed-specialist-teams, combining math- and verbal specialists, have comparable scores in math and verbal ability at the team level, but not at the individual level. Math-specialist-teams are composed of only math specialists and verbal-specialist-teams consist only of verbal specialists. Hence, these teams have a relatively high math or verbal score, both at the individual as well as at the team level average. We further make sure that the teams within one class are comparable in terms of average ability and gender composition. Otherwise the assignment of children to teams is random.[7]

### 4.2.2 Sample

The main source to measure math and verbal ability is a nationwide exam called the "CITO"-test. This exam consists of two mandatory parts, measuring math and verbal ability, respectively.[8] We received these CITO math and verbal scores from the schools for the majority of the pupils in our sample (i.e., 62%), see Table 4.1. The scores

our sample is described in more detail in Section 4.2.2.

[7]Per class we tried to form as many "usable" teams as possible, i.e., teams that fit the team composition shown in Figure 4.1. However, it also occurred that there were some children left that couldn't form a team fitting the rules. We placed these children together in one team and labeled this as "leftover"-teams.

[8]Besides, schools have the option to include a part testing 'world orientation', which is a combination of history, geography and biology.

**Table 4.1.** Overview of test score types

| **Test score type** | **Pupils** | | **Teams** | |
|---|---|---|---|---|
| | # | % | # | % |
| CITO *(standardized test score)* | 709 | 48 | 124 | 48 |
| CITO *(% of correct answers)* | 208 | 14 | 36 | 14 |
| LVS | 248 | 17 | 43 | 17 |
| DO | 110 | 7 | 18 | 7 |
| School grades | 57 | 4 | 10 | 4 |
| Unavailable | 155 | 10 | 25 | 10 |
| Total | 1487 | 100 | 256 | 100 |

were reported by the schools in two different formats: the official, standardized scores (48%) ranging from 0 to 100, where 50 corresponds to the nation wide average or, alternatively, the percentage of correct answers on both the math and verbal test per child (14%).

Not all schools in our sample participate in the CITO-test.[9] Almost all the schools that did not participate in this exam provided us either with the grades from a standardized student tracking system, called the "Leerlingvolgsysteem" (LVS) or with the scores from another type of nationwide exam, called the "Drempelonderzoek" (DO). The LVS records the pupil's progress from the first until the last grade of primary school. For each subject several standardized tests are conducted during each school year where test scores range from A to E. We use the math and verbal test scores from the LVS for 17% of the pupils in the sample. We obtained scores from the DO-test for 7% of the pupils in the sample. The DO-test is comparable to the LVS and CITO-test, with test scores ranging from 65 to 135, where 100 corresponds to the nationwide average. The scores from both the CITO and the DO-test can be converted directly into grades that correspond to the grades from the LVS. Finally, 10% of the schools (and pupils) did not provide any test scores (on time) and were removed from the sample. An overview of the number of pupils and teams per type of test score is shown in Table 4.1.[10]

Besides collecting test scores from the schools, we obtained information directly from pupils by means of two extensive questionnaires, one prior to the start of the program (pre-treatment) and one after the program (post-treatment). The overall response rate for both questionnaires is 92,5 %.[11] The pre-treatment questionnaire contains a wide variety of questions on individual background characteristics, such as age, gender,

[9] In the Netherlands approximately 80% of the primary schools participates in the math and verbal test of this exam.

[10] One school did not provide us with test scores from any of the above mentioned standardized tests, but gave us the students' grades (A-F) from a school exam instead.

[11] One class did not return the second questionnaire and for some children one of the questionnaires is missing due to absenteeism on one of the test days.

ethnicity, occupational status of the parents, etc. Additionally, both questionnaires contain some questions to assess the children's knowledge on entrepreneurship before and after the program. At the team level the post-treatment questionnaire is used to collect some information on team characteristics, such as the number of conflicts within a team and (self-assessed) teamwork. We have developed and tested these questionnaires in close collaboration with a child psychologist.

Data on team performance are obtained via the teachers. They filled out a standardized spreadsheet during and at the end of the program to register all transactions made by the teams, such as number of shares sold, share price, revenues etc. We use this spreadsheet to make sure that all the teachers collect the same information and calculate the team performance in exactly the same way.[12] The teams also registered all transactions themselves and completed a financial overview of their company at the end of each day. We consider the information collected by the teacher as providing the more objective and accurate results at the end of the program (teams turned out to sometimes make mistakes or miscalculations). This information is used to determine the winning team and the ranking of teams. The response rate for the completed excel files was 87,3% (i.e. for 8 out of 63 classes we did not receive the excel file). The final sample for which we have received all the required information, i.e. both the math- and verbal scores and the results on team performance, consists of 1131 pupils in 179 teams.

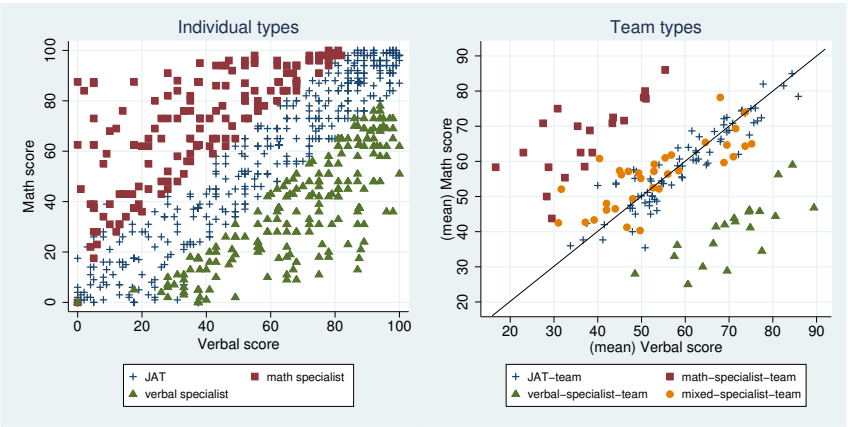
**Individual types**

A few weeks prior to the start of the education program we received the names of the children, their gender and their test scores from one of the tests described above. We measure the balance in mathematical and verbal ability by taking the (absolute) difference in the test scores between the two subjects. To define an individual as a JAT, we use a maximum difference of 15 (percentage) points as a cutoff point for the test scores from the nation wide exams (i.e. CITO- and DO-test). This cut-off point is chosen because it is the smallest unit of distinction for the grades in the LVS, i.e., a 15 point difference in test score corresponds to a one grade difference in the A-E scores of the LVS (and the grades from the school exam).[13] Specialists are defined as those children with a difference between the two test scores of more than 15 (percentage) points in the nationwide exam or a difference in grades of at least 1 for the LVS or the school exam. The use of this definition implies that the group of specialists also includes some pupils that are not very good in either of the two subjects. However,

[12] An example of the teacher spreadsheet is shown in Figure A4.1 in the Appendix.

[13] For the DO scores, we used an official conversion table to match the scores to the grades of the LVS to confirm that the classification of the individuals based on the DO scores is accurate (see: http://www.drempelonderzoek.nl).

**Figure 4.2.** Distribution of Individual and Team types




these specialists still have a sizable comparative advantage in one subject. Based on this choice of cut-off points, our sample consists of 720 JATs, 292 math-specialists and 306 verbal-specialists. The left hand side of Figure 4.2 shows the distribution of the individual types in the sample.[14]

### Team types

The individual types are used as the basis for our team composition (see Figure 4.1). Because participation in the education program is at the class level, the assignment of individuals to teams takes place within classes. Per class, the possibilities depend on the number of children (on average 24) and the distribution of individual types. Most teams consist of five or six pupils.[15] Two considerations guide the otherwise random allocation of children to teams within each class given their individual type: teams should be equal as much as possible in terms of average team ability and gender composition.[16]

After the categorization of the individual types in each class is done, the JATs are assigned to the JAT teams (see Figure 4.1) and then we (basically) either mix or separate the specialists (this choice is randomly distributed over classes). The resulting sample of teams consists of 117 JAT-teams, 41 mixed-specialist-teams, 23 math-

[14]Besides the 155 pupils for whom the test scores for the entire class are missing, 14 individual test scores are missing of pupils spread out over different classes. Hence, these children are assigned to "leftover"-teams.

[15]There are 18 teams with seven team members and one team of only four children.

[16]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.). Ability is defined here as the unweighted average of the sum of the two subjects. Otherwise none of the individual non-ability related characteristics was associated with team outcomes.

specialist-teams and 27 verbal-specialist-teams (see the right hand side of Figure 4.2). The other 23 teams were 'leftover' teams and consisted of combinations of individual types that couldn't be classified as any of the team types of interest.

The teachers were not informed about the details and the purpose of the team composition. We merely informed them about the resulting team compositions and that the teams should not be changed without our prior consent, unless they had strong objections against the team assignment. Based on the teachers' objections, 20 children moved teams prior to the start of the program. As a result, five teams were no longer usable. 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 of the usable teams. The final number of teams and the descriptives of some of the main characteristics per team type are shown in Panel A of Table 4.2.

To test for pre-treatment differences, we estimate whether team types differ in terms of their relevant average team background characteristics, see Table 4.2. As intended, given our experimental design, the teams differ in terms of average mathematical and verbal ability: math-specialist teams have a significantly higher math score and verbal-specialist teams have a significantly higher verbal score compared to the benchmark of JAT teams. Moreover, and in line with nation wide averages, girls in our sample score higher on verbal ability and boys score higher on mathematical ability. Therefore, despite our efforts to create balanced teams in terms of gender composition, the average share of females is significantly higher in verbal-specialist teams and significantly lower in math-specialist teams. In the same vein, even though our aim was to create teams of similar average ability, the (average) ability levels are significantly lower for mixed-specialist and verbal-specialist teams compared to JAT- and math-specialist teams. We will control for these differences in our estimations.

### 4.2.3 Outcome variables

For the evaluation of the effect of team composition on team performance, we use the following outcome measures. The first (and main) outcome measure is the team's ranking within their class. For the majority of the classes ($\frac{2}{3}$) the ranking of the team is based on the financial measure *Value of own shares*. This is calculated as total company value multiplied by the fraction of shares still owned by the team (i.e. not sold to the investor). For the remaining 33% of the teams in the sample the ranking equals the (unweighted) average of the ranking based on the *Value of own shares* and the ranking based on the number of *Sustainability Points*. The fact that there is no uniform performance measure underlying the team ranking is not problematic because the measure is uniform within each competitive environment, i.e. within one class.

On average there are 4.62 teams per class, with a minimum of two and a maximum

**Table 4.2.** Descriptives by team type

| | Total sample | | JAT-teams | | mixed-specialist-teams | | math-specialist-teams | | verbal-specialist-teams | |
|---|---|---|---|---|---|---|---|---|---|---|
| | mean | std. dev. | mean | std. dev. | mean | std. dev. | mean | std. dev. | mean | std. dev. |
| ***A: Background characteristics*** | | | | | | | | | | |
| Math score | 56.86 | 12.84 | 59.05 | 11.49 | 56.53 | 9.16 | 67.42** | 11.54 | 40.34*** | 9.22 |
| Verbal score | 57.13 | 13.82 | 59.67 | 11.30 | 53.09*** | 12.53 | 39.16*** | 11.18 | 69.48*** | 9.29 |
| $\|\Delta\|$ (Math - Verbal) | 10.41 | 11.58 | 3.28 | 3.00 | 6.55*** | 5.20 | 28.26*** | 6.94 | 29.14*** | 5.28 |
| % Female | 0.51 | 0.15 | 0.52 | 0.13 | 0.47* | 0.12 | 0.38*** | 0.16 | 0.65*** | 0.18 |
| Age | 11.73 | 0.33 | 11.72 | 0.33 | 11.70 | 0.31 | 11.85 | 0.37 | 11.73 | 0.36 |
| Average ability | 3.07 | 0.71 | 3.20 | 0.69 | 2.88*** | 0.76 | 3.09 | 0.64 | 2.76*** | 0.65 |
| Team size | 5.76 | 0.61 | 5.77 | 0.64 | 5.89 | 0.45 | 5.47** | 0.52 | 5.67 | 0.73 |
| Nationality parents: One non-dutch | 0.14 | 0.16 | 0.15 | 0.16 | 0.13 | 0.15 | 0.08** | 0.12 | 0.18 | 0.19 |
| ***B: Outcome variables*** | | | | | | | | | | |
| Normalized Team Rank | 0.57 | 0.27 | 0.53 | 0.27 | 0.64* | 0.28 | 0.49 | 0.19 | 0.72*** | 0.22 |
| Money won in Tournament | 1.23 | 1.30 | 1.33 | 1.35 | 1.08 | 1.26 | 1.73 | 1.16 | 0.62*** | 1.02 |
| Value of own shares | 109.35 | 100.95 | 122.31 | 111.57 | 91.28 | 81.49 | 130.57 | 104.41 | 60.38*** | 55.85 |
| Number of teams | 179 | | 105 | | 38 | | 15 | | 21 | |

*Note: */**/*** indicates a significant difference at the 10%/5%/1%-level in the team level average of the variable compared to the JAT-teams. Math- and verbal score and the absolute difference between the two scores are based on the nationwide exam (CITO) and the converted scores from the student tracking system (LVS). These are available for 150 teams. Average ability is proxied by 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). The values for Value of own shares are based on the sub-sample of teams where this was the measure to determine team ranking (N=113).

of nine teams per class.[17] Hence, team ranking can vary between one and nine, 1 being the best and 9 being the worst. Because it is easier to win in a class with a few teams than in a class with many teams, we divide the rank of each team by the number of teams in its class such that Team Rank is normalized between 0 and 1. We consider this *Normalized Team Rank* as our main outcome variable.

Additionally, we will use *Value of own shares* as an outcome measure for the sub sample of teams where this was the only measure to determine team ranking. This performance measure provides more detailed information on team performance than only ranking. The sub sample consists of 113 teams (61 JAT-teams, 25 mixed-specialist-teams, 14 math-specialist-teams and 13 verbal-specialist-teams) and is thus large enough to make useful estimations.[18]

As a final, alternative measure of team performance we will use the *Money won in the Tournament*. 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 (B) of Table 4.2.

### 4.2.4 Randomization

There are several factors that could affect the design of our field experiment and the internal validity of our findings in a negative way. One of the main concerns is non-random attrition due to a non-random selection of teams or classes that failed to report their results. To test for this, we regress an indicator for whether or not the team results were missing on the different team type dummies and on various background characteristics at the team level. The results from these estimations show that attrition was indeed random. The internal validity of our results further hinges on the random assignment of individual types to the various team types. This issue is only relevant for specialists who can be assigned to either mixed or unmixed teams of specialists. JAT-individuals, on the contrary, can only be assigned to JAT-teams. We tested whether

[17] There are just two classes with either of these two extreme values. More in general, there are eight classes with less than four teams per class and three classes with more than six teams.

[18] An overview of the parallel assignment to the team treatments and the sustainability treatments is shown in Table A2.1 in the appendix of Chapter 2.

the assignment of specialists to team types was random by running regressions at the individual level for both types of specialists separately. The dependent variable is a dummy variable indicating whether an individual is assigned to a mixed-specialist-team (value 1) or to a specialist team matching his/her own type (value 0), whereas the independent variables are individual characteristics such as age or gender. The estimation results in Table A4.1 in the Appendix indicate that the assignment was indeed random.[19]

Finally, we check whether the changes to the team composition that were made on the teacher's request after the initial assignment (but prior to the start of the program) affect our research design. We do this by comparing the (observed) team characteristics of the initial teams with the teams in our final sample. The results from this comparison show no systematic (significant) differences.

## 4.3 Results

### 4.3.1 Main results

To estimate the effect of the different team types on team performance we regress each of the outcome variables on the team type dummies, using the JAT-teams (i.e. the largest group in our sample) as the benchmark. We control for several team and class level characteristics.[20] The team level characteristics we include are average age, gender composition (i.e. the share of females in the team), team size and average team ability. At the class level we control for the total number of teams per team type per class. The rationale behind these class level controls is that the team composition as well as the competition between teams took place at the class level. As a result, in some classes all the different team types are represented in the tournament, whereas in other classes certain team types are missing or might be overrepresented. Thus, by including the sum of teams of each team type per class, we account for the differences in the tournament composition across classes. The observations are clustered at the class level to obtain robust standard errors to account for the possibly correlated performance of the teams within one class.

To answer the research questions, we test two hypotheses. For the first question we test whether JAT-teams perform better than teams consisting of individual spe-

[19]Two differences are (only) marginally significant. Mixed-specialist teams have slightly fewer female math specialists and math-specialists have slightly higher math scores than those in pure math-specialist teams. Given our effort to compose balanced teams in terms of gender and ability, none of these two differences are surprising because verbal specialists are more likely to be female and of a slightly lower ability level (see Table 4.2).

[20]The results for specifications without these controls are similar though, and can be found in Table A4.2 of the Appendix.

**Table 4.3.** The effect of team type on team performance

| | Normalized Team Rank | | Value of own shares | | Money won in Tournament | |
|---|---|---|---|---|---|---|
| | (1) | | (2) | | (3) | |
| math-specialist-team | 0.01 | (0.07) | -4.20 | (27.90) | 0.23 | (0.37) |
| mixed-specialist-team | -0.12* | (0.07) | -50.48** | (23.98) | -0.29 | (0.28) |
| verbal-specialist-team | -0.23*** | (0.07) | -60.19*** | (19.58) | -1.21*** | (0.31) |
| *(omitted category: JAT-team)* | | | | | | |
| **Team characteristics** | | | | | | |
| Age | 0.17*** | (0.05) | 41.28 | (29.59) | 0.59*** | (0.22) |
| Female | 0.03 | (0.15) | -125.85* | (72.89) | 0.53 | (0.64) |
| Average ability | 0.01 | (0.02) | 14.54 | (18.33) | -0.04 | (0.11) |
| Team size | 0.06 | (0.03) | 24.22 | (20.82) | 0.15 | (0.16) |
| **Class characteristics** | | | | | | |
| *# of teams per class:* | | | | | | |
| JAT-teams | 0.002 | (0.01) | -20.79 | (16.89) | -0.30*** | (0.05) |
| math-specialist-teams | 0.03 | (0.03) | -12.50 | (27.79) | -0.26* | (0.14) |
| verbal-specialist-teams | 0.07*** | (0.02) | 23.18 | (16.15) | 0.15 | (0.10) |
| mixed-specialist-teams | 0.03 | (0.02) | 14.17 | (17.00) | -0.20** | (0.08) |
| leftover-teams | 0.02 | (0.02) | -6.31 | (21.54) | -0.16 | (0.10) |
| Constant | -2.93*** | (0.68) | -447.53 | (474.49) | -5.47* | (2.91) |
| Number of teams | 178 | | 112 | | 178 | |
| $R^2$ | 0.14 | | 0.17 | | 0.17 | |

* Note: */**/*** indicates a significance at the 10%/5%/1%-level. Observations are clustered at the class level. Robust standard errors in parentheses. Average ability is proxied by the intended future high school track and ranges from 1 (pre-vocational secondary education) to 5 (pre-university education). To simplify the interpretation of the estimated coefficients for Normalized Team Rank, we multiply these values by -1. The results for Value of own shares were obtained using a sub sample for which this outcome measure was the only determinant for team ranking.

cialists do. To answer the second question we compare the performance of JAT-teams and mixed-specialist teams. In both cases we test the null hypothesis that there is no difference in performance. The estimation results are shown in Table 4.3. To simplify the interpretation of the estimated coefficients for Normalized Team Rank, we multiply these values by -1. The estimated coefficients indicate the difference in performance of each team type compared to the JAT-teams (with robust standard errors in parentheses). Each column shows the result for one of the (three) outcome variables. When comparing the results in the three columns in Table 4.3, a consistent pattern emerges. JAT-teams and math-specialist teams perform equally well in terms of all three outcome variables, i.e. *Normalized Team Rank*, *Value of own shares* and *Money won in the Tournament*. This finding is partly in line with our first hypothesis that JAT-teams would outperform teams consisting of one type of individual specialists. The hypothesis holds for the verbal-specialist teams but not for the math-specialist teams.

Furthermore, we find that the performance of the JAT-teams is significantly better than the performance of the mixed specialists. These results indicate that the ability to combine resources effectively is not something that comes across when people combine their specialized skills within teams. Note that we obtain a negative coefficient for the mixed-specialist-teams for all three outcome variables. For two of the outcome variables, i.e. *Normalized Team Rank* and *Value of own shares*, this difference is statistically significant ($p$-values: 0.09 and 0.04). For the third outcome measure, *Money won in the Tournament*, the coefficient is not significantly different from zero. Thus, the performance of the mixed-specialist-teams is on average lower than the performance of the JAT-teams. This indicates that it is not possible to compensate the lack of skill balance at the individual level by combining (the skills of) two types of specialists within one team.

Finally, the setup of the experiment allows us to analyze the effect of the two specific skills, i.e. math and verbal ability, separately. As stated above, math-specialist-teams perform equally well as JAT-teams. Thus, in line with the findings by Hartog et al. (2010) for solo entrepreneurs, we find that mathematical ability is also beneficial for entrepreneurial teams (some additional evidence is found in Van Praag and Cramer, 2001). The equal performance of JAT-teams and math-specialist teams suggests some substitutability between mathematical ability and balanced skills. Math-specialist teams have significantly higher mathematical ability than the JAT-teams (see Table 4.2). Yet, JAT teams are able to make up for this lack in mathematical ability by the presence of individuals with more balanced skills within their team. Alternatively, one could argue that math-specialist teams compensate the missing skill balance by sufficient mathematical ability. Nevertheless, it appears that team composition is important. Mixed-specialist teams have the same mathematical ability and the same

skill balance at the team level as the JAT-teams, but show lower performance. Thus, the benefits of skill balance only exist if a team consists of only JATs. Moreover, if we control for average math ability and the highest math score within the team, the coefficients of these variables are insignificant and the results remain the same (albeit slightly less significant due to the loss of observations). This indicates that the results are not driven by just one high ability math person or superstar.

The effects of verbal ability are very consistent, and (relatively) negative, for all three outcome variables. Table 4.3 shows that the performance of the verbal-specialist-teams is significantly lower compared to the JAT-teams for all three outcome measures. This is also consistent with the findings by Hartog et al. (2010) that verbal ability does not yield a positive income effect for (individual) entrepreneurs.

The class or tournament characteristics provide some (additional) support for our findings. The estimations show that a larger number of JAT-teams or math-specialist-teams in the class is negatively related to team performance in terms of *Value of own shares* and *Money won in the Tournament*. Additionally, we find that a larger number of verbal-specialist-teams in the class is positively related to (all three) outcome measures (although only significant for *Normalized Team Rank*). This is exactly what one would expect based on the main effects: JAT teams as well as teams consisting of math specialists provide fiercer competition because they outperform the others.

### 4.3.2 Mechanisms

Given the results we find, it would be valuable to be able to address the issue of *why* teams with JATs and math specialists perform better than teams of mixed or verbal specialists.

The literature reports both advantages and disadvantages of team diversity for team performance (Hamilton et al., 2003; Hoogendoorn et al., 2013; Kilduff et al., 2000). Possible gains from diversity could occur through complementarities and mutual learning (Lazear, 1999; Dahlin et al., 2005). However, these benefits might be offset by costs associated with communication and coordination which becomes more cumbersome in a diverse team (Lazear, 1999; Richard et al., 2004; van Dijk et al., 2012). We can measure coordination and communication costs in terms of the number of conflicts in a team and in terms of teamwork, where the latter is an inverse measure of these costs (similar to Hoogendoorn and Van Praag, 2012). These are based on two team variables measured in the post-treatment questionnaire for this purpose: the number of conflicts within a team and teamwork. Both measures are self-assessed. The number of conflicts is measured in terms of the average of the number of reported conflicts by individual team members within one team. Teamwork is measured by the (team average) answer to the question: "How well did you work together?". We measure this on a 5-point

scale where 1 = very well and 5 = very bad.

Comparing the average number of conflicts per team across team types, we see that there are indeed more conflicts in the mixed-specialist teams than in the (more homogeneous) JAT-teams. However, this difference is not statistically significant. Teamwork, our second measure, might also be easier accomplished in less diverse teams. The differences in teamwork across team types show that mixed-specialist teams (and verbal-specialist) teams score (marginally) significantly lower compared to JAT-teams ($p$-value: 0.08).[21] Hence, we find some evidence of higher coordination and communication costs in teams of mixed specialists compared to JATs.

These costs may potentially explain the better performance of JAT teams compared to mixed specialists if we also find a direct effect of these costs on performance. However, the design of our experiment does not allow any causal inferences other than related to team type and team performance due to a lack of other exogenous variation.[22] For instance, the self-reported measures of conflicts and teamwork might be affected by the performance of the teams, rather than the other way around, where the winning teams judge their teamwork more positively ex-post than the members of the losing teams. A factor that limits the applicability of this explanation is that the (homogeneous) teams of verbal specialists also have lower scores on teamwork than JAT and math-specialist teams. This is at odds with the explanation that diversity would cause higher coordination and communication costs.

Our final exercise to find suggestive evidence of what underlying mechanism might explain our results is related to the benefits side of more diverse teams. More diverse teams would have more complementary skills (which we imposed on the mixed-specialist teams but not on the others) and this might result in better (mutual) learning outcomes. To test this assertion we regress the development of entrepreneurship knowledge and nine non-cognitive skills typically associated with entrepreneurship on the team type dummies (see Table 3.3 for an overview of the non-cognitive skills). We do not find any differences in skill development across the different team types.

All in all these results can only provide suggestive evidence about the association between team *diversity* and performance. The distribution over team types of the self-assessed numbers of conflicts within the team and the quality of the teamwork suggests that some costs of diversity are associated with mixed-specialist teams. However, we find the same disadvantage for the relatively homogeneous teams of verbal specialists. Therefore, we conclude that it is somewhat unlikely that the greater *diversity* of the

[21] Once we control for team characteristics such as team size, gender composition and average ability, the coefficients remain negative but they are no longer significant for mixed-specialist teams.

[22] We find an insignificant relation between the average number of conflicts in a team and performance and a significant and positive relation between teamwork and performance (according to all three performance measures).

teams of mixed specialists causes their lower performance relative to JATs.

### 4.3.3 Robustness checks

We have performed several checks to confirm the robustness of our findings. Firstly, the results are robust to the in- or exclusion of team and class characteristics as control variables. Secondly, we ran separate regressions excluding the teams from our sample that were changed, upon the teacher's request, after the initial assignment. Excluding the five affected teams from our estimation sample in which an individual team member was replaced by another, but where the team type remained unchanged, leaves the results qualitatively the same.

To make sure that the results we find are not driven by the four JAT-teams with very high math (and overall) ability (see Figure 4.2), we exclude those four teams from the sample. The results from these estimations are similar to the main results presented above, i.e. the performance of mixed-specialist teams compared to JAT-teams is lower in all three specification, but is only (marginally) significant for the outcome measure *Value of own shares*.

In another robustness check we test whether the results are not driven by a few superstars within the teams. To do this we create dummy variables of being a superstar for those individuals at the top 10% of the ability distribution for verbal and mathematical ability respectively. More specifically, since mathematical ability appears to be important at the team level, we test if performance is different for teams with a *math superstar* by including an interaction term of the *math superstar* dummy and the different team types. The coefficients for these interaction terms are insignificant in all specifications, so there appears to be no additional benefits from having a *math superstar* in the team.

An alternative explanation for the high performance of the JAT-teams compared to mixed-specialist teams is that, by definition, only JAT-teams include pupils who score very high on both math and verbal ability. *JAT superstars* are individuals whose test scores for both math and verbal ability are in the top 10% of the distribution (about half the JAT-teams has a *JAT superstar* according to this definition). If the performance of the JAT-teams is driven by only a few superstars, then our findings are not the results of teamwork, but rather the result of one high ability JAT person leading the team to high performance. To test if having a *JAT superstar* affects the performance of the JAT-teams, we use the subsample of JAT-teams and regress team performance on a dummy variable indicating if a team has one (or more) JAT superstars in the team, including the other controls we use in the main estimation. The coefficient for this dummy is not significant in any of the specifications, indicating that the performance of the JAT-teams is not driven by just one superstar in the team.

Finally, we estimate our main specification(s) without the teams that do not provide us with the (official) test scores from the nationwide exam or the standardized student tracking system. That is, we excluded those classes for which we received less precise measures, i.e., the percentage of correct answers from the nationwide exam or the grades of school exams (see Table 4.1). The results are very similar, although somewhat less significant due to the loss of observations.

## 4.4 Summary and conclusion

Teamwork is a growing phenomenon within firms of all types and sizes (Hamilton et al., 2003). Furthermore, the tasks within organizations (from start-ups to large firms) have become increasingly complex (Dahlin et al., 2005; Lazear and Shaw, 2007). The combination of an increase in task complexity and the observed increase in teamwork raises the question of effective team composition. In this chapter, we empirically explore this question focusing on the *skill* composition of successful teams. This chapter is partly motivated by the findings presented by Rulke and Galaskiewicz (2000) and it aims to answer two specific research questions. The first question is whether teams consisting of a combination of individual JATs show better performance than teams consisting of individual specialists do. The second research question builds on the study by Rulke and Galaskiewicz (2000) and is motivated by one of the predictions made by Lazear (2005). For solo entrepreneurs skill balance, or being a Jack-of-All-Trades (JAT), appears to be beneficial. However, Lazear (2005) predicts that as production processes become more complex, it will become more and more difficult to find all the required skills to be successful as an entrepreneur endowed within one person. Hence, the second research question investigates whether it is possible to substitute (a lack of) individual balanced skills by combining the skills of different specialists within one team.

To answer these questions we have conducted a field experiment. We find an environment to analyze the role and substitutability of balanced skills for teams with entrepreneurial tasks within the rather unusual setting of an entrepreneurship education program ("BizWorld") in primary schools in the Netherlands. Teams of 5-6 children in the last grade set up a toy business in friendship bracelets. Two skills are relevant at this age: verbal and mathematical ability. Based on uniform and valid measures of the scores on these skills, we compose four different team types: JAT-teams, verbal-specialist teams, math-specialist teams and mixed-specialist teams. These teams set up a business in a competitive environment and we obtain uniform measures of the teams' performance.

Comparing the performance of the different team types we find that JAT teams

do (weakly) better than teams consisting of individual specialists. In particular, JAT-teams perform significantly better verbal-specialist teams. Math-specialist teams perform equally well. Taken together these results indicate that balanced skills are indeed beneficial for the performance of teams within this dynamic (entrepreneurial) environment. Moreover, the results show that JAT-teams perform significantly better than mixed-specialist teams. The lower performance of the mixed specialist teams indicates that it is hard to substitute individual balanced skills by combining different specialists within one team. However, more research is needed to determine the exact mechanism through which this lower performance occurs.

Obviously, this experimental design has some limitations. The most important one is the possibly limited external validity of our results. The fact that we study teams of children participating in an entrepreneurship education program instead of actual entrepreneurial teams, poses limitations to the generalizability of our results. However, (a priori) there is no clear reason to assume that the treatment effects found for children should vary substantially with age or subject pool. Another possible limitation is that we define JATs and specialists only in terms of two types of skills: mathematical and verbal ability. As we noted already, according to the psychological literature, these are considered the two core skills for children at the age of 12 and form the basis of relevant skills to be developed later in life.

Besides these limitations there are some notable advantages of our field experiment which allows us to measure the performance effect of balanced skills with a high level of internal validity. Most importantly, there is no self-selection bias, neither into or out of working together in a team in general, nor of the specific team members. So far, few empirical studies on effective team composition have taken into account that team composition is non-random and that the performance effects that are measured might be biased as a consequence. The studies that do allow causal inferences have not taken place in environments requiring a broad array of skills and involving complex problem solving and decision making, which is a realistic feature of influential teams, such as management boards or teams of entrepreneurs. The entrepreneurship education program provides us with a suitable real effort team task that arguably has these characteristics. Thus, this study provides an important first step in unraveling the effective skill composition of (entrepreneurial) teams.

# Appendix

Table A4.1: Random assignment of individual specialists to teams

| | **Verbal specialists** | | **Math specialists** | |
|---|---|---|---|---|
| | *Mixed-specialist-team dummy* | | *Mixed-specialist-team dummy* | |
| Age | -0.06 | (0.11) | -0.13 | (0.11) |
| Female | 0.06 | (0.06) | -0.09* | (0.05) |
| Math score | -3.80 | (2.93) | 5.93* | (3.23) |
| Verbal score | -3.08 | (3.47) | 0.66 | (3.55) |
| High school track (1-5) | -0.01 | (0.19) | -0.14 | (0.22) |
| Nationality parents: both non-Dutch | 0.10 | (0.09) | -0.002 | (0.09) |
| Entrepreneurship knowledge | -0.25 | (0.22) | -0.38 | (0.23) |
| Number of teams | 227 | | 200 | |

*Note: The coefficients in each cell come from separate regressions of the row variable on a dummy variable that takes the value 1 if an individual was assigned to a mixed-specialist-team and 0 if an individual was assigned to a pure specialist team matching his own type. Observations are clustered at the class level, robust standard errors in parentheses. Math- and verbal score are based on the nationwide exam and the converted scores from the student tracking system, e available for 182 math specialists and 181 verbal specialists. High school track corresponds to the intended future high school track and ranges from 1 (pre-vocational secondary education) to 5 (pre-university education). Entrepreneurship knowledge is based on the number of correct answers to eight questions about entrepreneurship. */**/*** indicates significance at the 10%/5%/1%-level.

Table A4.2: The effect of team type on team performance (without controls)

| | Normalized Team Rank (1) | | Value of own shares (2) | | Money won in Tournament (3) | |
|---|---|---|---|---|---|---|
| Math-specialist-team | 0.04 | (0.06) | 8.26 | (26.49) | 0.40 | (0.36) |
| Mixed-specialist-team | -0.11* | (0.06) | -31.03 | (20.43) | -0.25 | (0.26) |
| Verbal-specialist-team | -0.19*** | (0.05) | -61.93*** | (20.00) | -0.71*** | (0.23) |
| *(omitted category: JAT-team)* | | | | | | |
| Constant | -0.53*** | (0.02) | 122.31*** | (15.37) | 1.33*** | (0.11) |
| Number of teams | 179 | | 113 | | 179 | |
| $R^2$ | 0.07 | | 0.05 | | 0.05 | |

* Note: */**/*** indicates a significance at the 10%/5%/1%-level. Observations are clustered at the class level. Robust standard errors in parentheses. The results for Value of own shares were obtained using a sub sample for which this outcome measure was the only determinant for team ranking.

Figure A4.1: 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 spent** | | | **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 | # own shares | **Team company value** | number of team members |
| Friends & Co. | 269 1 | 104 | 7 | 1 | 10 | 86 | 97 | 60 | 37 | 11 | 26 | 7 | 182 | 6 |
| V and B | 269 2 | 59 | 7 | 1 | 10 | 41 | 55 | 61 | -6 | 0 | 0 | 6 | 0 | 6 |
| Made by Uzz! | 269 3 | 84 | 7 | 1 | 10 | 66 | 82 | 63 | 19 | 5 | 14 | 7 | 98 | 6 |

# Chapter 5

# Incentives and sustainable behavior

*This chapter is based on Huber, L.R., R. Sloof, and M. van Praag, The effect of incentives on sustainable behavior: Evidence from a field experiment.*

## 5.1 Introduction

Nowadays, there are examples abound of all kinds of pro-social behavior. For example, at the corporate level, many companies include CSR into their mission statement or corporate strategy. Firms respond to customer demands and offer more sustainable products. Also in the labor market, the social image of a future employer seems to be of increasing importance (Kitzmueller and Shimshack, 2012).[1] At the individual level people also engage in (pro)social behavior, ranging from blood donations, buying fair trade products to choosing socially responsible banks and investment funds (Bénabou and Tirole, 2010). As such, pro-social behavior (both at the individual and at the corporate level) is a topic of interest to practitioners and researchers alike. Broadly speaking, two important questions have been raised: (1) How can socially responsible behavior be induced? and, (2) Is socially responsible behavior associated with financial performance?

In this study we focus on the first question that has been identified in the more recent literature as one of the core questions related to CSR (see Kitzmueller and Shimshack (2012)).[2] Our paper is motivated by two strands of literature. First, a set of recent field experiments shows that financial and non-financial incentives have a positive effect on social behavior (Ashraf et al., 2014; Miller et al., 2012; Olken

[1] Several definitions of CSR exist. Kitzmueller and Shimshack (2012) define CSR as follows: "CSR is corporate social or environmental behavior that goes beyond the legal or regulatory requirements of the relevant market(s) and/or economy(s)" (p.53).

[2] Kitzmueller and Shimshack (2012) also include an extensive analysis on the somewhat less recent literature that focused on the question of the legitimacy of CSR, that we take for granted in this study.

et al., 2012). These experiments are all conducted in settings in which social behavior (related to health and educational outcomes) is the main aim of the program and thus the main outcome variable of interest. However, social behavior has also become a topic of interest for many (large) for-profit firms, and social goals are sometimes added to traditional financial goals (Kitzmueller and Shimshack, 2012). Within these companies a method has to be found to effectively balance these two (sometimes conflicting) goals. Besley and Ghatak (2013) develop a selection model as a potential solution to effectively balance profit with purpose. However, they show that the supply of motivated agents that are needed to successfully manage such hybrid organizations might not be sufficient. An alternative solution to induce sustainable behavior could be through the use of incentives, the topic of our study.

Various studies have looked at the association between incentives and corporate social performance, and the empirical evidence is rather mixed (see Walls et al. (2012) for an overview). The drawback of many empirical studies is that the matching between a CEO and a firm, and between a firm and the remuneration policy, is likely to be endogenous. Hence, it is virtually impossible to establish a causal link, when studying the association between CEO incentives and CSR using only observational data.

To solve these methodological difficulties and to answer our research question we employ an experimental design to study the effect of incentives on social behavior in a setting in which financial performance also matters. As described above, social behavior can take on many forms. In this paper we focus on sustainable (or environmental) behavior. We conduct a field experiment within the context of an entrepreneurship education program and study the behavior of the entrepreneurial teams of children (aged 11 or 12) in response to three different treatments (baseline, non-incentivized CSR and incentivized CSR).[3] The baseline consists of the regular program, and thus serves as our control group. We add two specific elements to the regular program (also for the control group) to reflect sustainability, i.e., sustainable yarn and a sustainability trademark. In the two CSR treatments more emphasis is placed on sustainability by explicitly discussing sustainable behavior and the importance of the environment.[4] In the non-incentivized CSR treatment the reward structure is the same as in the baseline treatment and only depends on the financial performance of the team. In the incentivized CSR treatment the performance of the teams is evaluated based on a combination of financial performance and sustainable behavior.

The results indicate that there is no difference in terms of sustainable behavior

[3]Using the classification defined by Harrison and List (2004), one could argue that this is a framed field experiment rather than a natural field experiment since the education program is quite different from the regular course content in primary schools.

[4]As stated above, we define social behavior in this paper in terms of sustainable behavior. However, we use the term CSR to label our treatments for the ease of abbreviation.

between the teams in the baseline treatment and the teams in the non-incentivized CSR treatment. Thus, merely discussing the environment and emphasizing its importance, does not cause a behavioral change. The estimates from the comparison between the two CSR treatments (i.e., with and without incentives) show that incentives have a significant positive effect on sustainable behavior. Moreover, descriptive evidence suggests that the two CSR treatments also (positively) influence the children's attitude towards the environment. Since a trade-off is often assumed if financial and social goals are combined (within one organization), we also estimate the effect of the different treatments on financial performance. The results show that there are no significant differences across the three treatments in either of the financial outcome measures. This implies that, in this setting, the choice to behave more sustainable does not affect financial performance.

This study aims at contributing to the existing literature in several ways. Motivated by the field experiments on social behavior and the more descriptive evidence on the association between incentives and CSR, we combine the two topics and use an experimental design to study sustainable behavior in a productive environment in which both sustainable behavior and financial performance matter. Furthermore, within this productive setting we look at sustainable behavior at the team level, instead of at the individual level. This is a realistic feature, since production decisions are often made by teams. Incentives might prove to be a more effective way to induce sustainable behavior in teams (rather than selection of motivated individuals that then need to be combined in a team in specific ways). An obvious limitation of our study is that the experimental setting that we exploit (i.e., the education program) is very stylized, and thereby limits the external validity of the results.

Our experimental design has some benefits that lead to a high internal validity. First and foremost, due to the randomized treatment assignment we are able to estimate a causal effect of incentives on sustainable behavior. Second, conducting the experiment among children allows us to study a large sample and to provide strong financial incentives. Third, all teams operate within the same (simplified) environment and industry. This means that we have objective and uniform outcome measures for all the teams in the sample. Finally, within this setting the benefits of a real effort experiment are preserved. Teams work together for several days and have strong incentives to care about their product and the performance of their team.

Besides the methodological benefits of this experimental setting, studying sustainable behavior of children could be an interesting starting point to study the development of attitudes towards the environment more in general. A recent strand of literature started by Cunha and Heckman (2007) suggests that teaching certain skills at a young age might produce positive spillover effects in later periods. The findings on the envi-

ronmental attitude of the children in the two CSR treatments (measured three to eight weeks after the program) potentially suggest that encouraging and inducing sustainable behavior in primary schools might lead to positive attitudes towards the environment later in life.

The paper is structured as follows. An overview of the related literature is provided in Section 5.2. In Section 5.3 the design of the field experiment is explained in more detail. The results are shown in Section 5.4 and in Section 5.5 we put our results into perspective and conclude.

## 5.2 Related literature

The research question addressed in this paper is motivated by several recent papers related to individual and corporate social behavior. This section describes the studies that we alluded to in the introduction in more detail.

There is ample empirical evidence that monetary incentives are an effective tool to induce productive effort provision (e.g. Lazear, 2000; Bandiera et al., 2011). However, before turning to the specific setting of providing incentives for socially responsible behavior, we need to understand the potential drivers underlying this type of behavior more in general. In the economic literature three individual motives for pro-social behavior are described: intrinsic motivation, extrinsic motivation and social norms or image concerns.[5] Theoretical predictions and empirical results indicate that these motivations do not stand alone. There are several studies that look at how pro-social behavior is influenced by intrinsic and extrinsic motivations (e.g. Frey and Oberholzer-Gee, 1997; Gneezy and Rustichini, 2000a,b) and how these two types of motivations interact in the presence of social norms or image concerns (e.g. Ariely et al., 2009; Bénabou and Tirole, 2006). Similar to the theoretical predictions on individual pro-social behavior, the motivations underlying socially responsible behavior at the corporate level can be divided into social and classical preferences (Kitzmueller and Shimshack, 2012).[6] The type of CSR that is adopted by a company depends on the interaction between the preferences of shareholders and managers and this in turn influences the predicted association between CSR and financial performance.[7]

[5]Intrinsic motivation is related to the desire to do good (i.e. altruism or the value of giving per se). Extrinsic motivation is about providing external rewards in order to induce a certain behavior (e.g. thank-you gifts or tax benefits). Image concerns provide a motivation that is partly driven by the perception of others (Bénabou and Tirole, 2006; Ariely et al., 2009; Bénabou and Tirole, 2010).

[6]Social preferences refer to intrinsic motivation, and classical preferences refer to monetary incentives or extrinsic motivation.

[7]Bénabou and Tirole (2010) distinguish three types of motivations for firms to engage in CSR. The "long-term perspective" uses CSR to reduce the short-term view of managers in order to maximize profits. In case of "delegated philanthropy" the stakeholders use the firm to express their values, e.g. by buying their coffee at Starbucks because they sell fair-trade coffee. Finally, "insider initiated

To test how profit and social purposes can be successfully incorporated into one firm, Besley and Ghatak (2013) develop a selection model in which social enterprises select managers based on their motivation for social goals. They theoretically show and experimentally confirm that it is possible to use selection, i.e. the hiring of motivated agents, as a tool to create these so-called hybrid organizations. However, they find that the motivated agents that are needed for such organizations are scarce. These findings suggest that selection might not be the best, or at least not the only, way to incorporate financial and social goals within one company. Another way to induce socially responsible behavior might be by using incentives, which is the focus of our study.

A set of recent field experiments study the effect of (financial) incentives on social behavior (e.g. Ashraf et al., 2014; Miller et al., 2012; Olken et al., 2012). Ashraf et al. (2014) conduct a field experiment to estimate the effect of monetary and non-monetary incentives in a public health organization in Zambia and how these incentives correlate with intrinsic motivation. The study includes a control treatment, two financial treatments (with high and low financial margins) and a non-financial treatment. They find that non-financial incentives have the strongest effect on their outcome measure (i.e., condom sales). Moreover, they find that in their sample there is no evidence that incentives crowd out motivation. On the contrary, their results show that motivated agents (measured by a donation above the median to a HIV/AIDS charity) respond more strongly to both financial and non-financial incentives. The field experiment conducted by Olken et al. (2012) estimates the effect of financial incentives on the effectiveness of aid programs targeted at maternal and child health and education. To estimate the effect of performance incentives the researchers use an incentivized and a non-incentivized version of the program as well as a control group.[8] The results show that the incentives increased the performance based on health indicators but that they had no effect on the education indicators. Finally, the paper by Miller et al. (2012) reports the results of a randomized trial in primary schools in rural China. The purpose of this experiment is to estimate the effect of incentives on the reduction of anaemia. The experiment consists of three different treatments (information, subsidy and incentives) and a control group.[9] The comparison between these three treatments and the control

corporate philanthropy" is driven by the managers' personal preferences (for example for a certain charity).

[8]The program in the incentivized and the non-incentivized treatment is identical in terms of money, target indicators, information and monitoring tools.

[9]In all three treatments the school principal receives information on: the share of anaemic students in his/her school, effective methods to reduce iron deficiency anaemia and the relationship between anaemia and educational outcomes. In the "subsidy" treatment the principals received the information plus some specific operating budget to realize the program's objective. The "incentive" treatment consists of the information, the subsidy and performance payments if anaemia was reduced among their student population.

group shows that only in the incentive treatment there is a significant increase in the haemoglobin concentration of the student population and an associated reduction in anaemia compared to the control group. They do not find any differences among the three treatments. Taken together the results from these three field experiments suggest that incentives may be a useful tools to increase social behavior. However, all the programs described in the studies above look at situations in which individual social behavior is the main outcome variable of interest. In our study we want to estimate the effect of incentives on sustainable behavior in a setting where both financial and social goals matter for team performance.

Various studies have measured the relationship between CEO incentives and Corporate Social Performance (CSP) (see Walls et al. (2012) and Kitzmueller and Shimshack (2012) for an overview), albeit none in an experimental setting. These studies have established mixed evidence on the link between corporate governance and CSP. For example, Deckop et al. (2006) find a negative association between CSR and the bonus share as part of the overall compensation and McGuire et al. (2003) find no association between short-term incentives and either good or bad social performance. For long term incentives the results are also mixed. The findings by Coombs and Gilley (2005) suggest that there is no association between stock options and any CSR dimension and McGuire et al. (2003) find a positive relationship between long-term incentives and poor social performance (i.e., to avoid negative influences from bad social behavior). Moreover, the studies by Deckop et al. (2006) and Mahoney and Thorne (2005) also show positive associations between long-term incentives and CSP. Additionally, going back to the paper by Besley and Ghatak (2013), a trade-off is often assumed between meeting financial goals and CSR purposes (Kitzmueller and Shimshack, 2012). However, the evidence on the correlation between corporate social and financial performance is rather mixed, and slightly more often positive than negative (e.g. Garcia-Castro et al., 2010; McWilliams and Siegel, 2000; Waddock and Graves, 1997).[10] Overall, the mixed findings on the association between CEO incentives and CSR and between CSR and financial performance may indicate that it is difficult for CEOs to focus on two (sometimes conflicting) organizational outcomes. However, cause and effect remain unidentified in these studies due to the endogenous nature of the choice for these incentives.

[10]In an extensive meta-analysis, Margolis et al. (2007) show a very modest positive relationship between the two. They find that the median correlation coefficient is only 0.08 and that it decreases significantly when including basic controls such as industry and firm size. Moreover, the authors note that the direction of the causation could run both ways (from CSR to CFP and the other way around).

**Table 5.1.** Treatment options

| Treatment | Profit incentives | CSR discussed and encouraged | CSR incentives |
|---|---|---|---|
| Baseline | √ | | |
| non-incentivized CSR | √ | √ | |
| incentivized CSR | √ | √ | √ |

## 5.3 Data and methodology

### 5.3.1 Design of the field experiment

The objective of this field experiment is to examine the effect of incentives on sustainable behavior. More specifically we look at the differences in sustainable behavior and financial performance between three different treatments, i.e. the baseline treatment, the non-incentivized and the incentivized CSR treatment (see Table 5.1).

In the *baseline* treatment the pupils participate in the "regular" program with the addition of sustainable yarn and the sustainability trademark (see Section 2.1.1 for a detailed description). In the two CSR treatments sustainability is introduced and discussed at the start of the second day (see Table 2.1). The children are stimulated to think about how their team could contribute to a better environment, for example by reducing waste or by reusing the left over materials (for example by creating other products instead of bracelets). Moreover, the children in these two treatments are explicitly encouraged to include CSR into the mission statement. The difference between the two CSR treatments is in the incentive provision.

The reward structure (i.e. the performance measures that will determine the winning team) and the prizes are communicated to the children at the start of the program. In the *baseline* and in the *non-incentivized CSR* treatment the performance of the teams is assessed by means of a financial performance measure. The performance measure that is used to determine the ranking of the team within the class is the profit of the team multiplied by the number of shares still owned by the team. In the incentivized CSR treatment the performance of the teams is based on a composite performance measure including both financial and sustainable performance indicators. In this treatment the final performance of the teams is determined by the unweighted average of the team's ranking based on their financial performance and the rank based the team's sustainable behavior. Sustainable behavior is measured in terms of the percentage of sustainable yarn, and two dummy variables: one indicating whether or not a team included CSR into their mission statement and one for the use of plastic packaging material (see Section 5.3.3 for a detailed description).

The timing of the field experiment is as follows.[11] Prior to the start of the ex-

[11]The description of the timing of the events in this paper is very similar to that reported in Chapter

periment, all the children in the sample have to complete a questionnaire to collect some background characteristics of the individual pupils, e.g., age, gender, ethnic background, etc. This questionnaire, accompanied by a letter including information for the parents about the research project, is sent out to all schools in the sample in February.

During a train-the-trainer session that the teachers and entrepreneurs attended prior to the start of the program the course material, containing all the details about the education program, is handed out. The guidelines for the program are very strict and described in detail in an instruction manual included in the course material. The particulars of the different treatments are also described in this course handbook. During the training, the teacher and the entrepreneur of each class receive a hand out sheet stating the assigned treatment including a reference to the pages in the handbook where the details of their treatment are mentioned. After the training an e-mail was sent to both the entrepreneur and the teacher, containing the same information as in the hand out, to confirm the assigned treatment. It is emphasized, both during the training and in the e-mail, that they should not deviate from the course content (including the treatment information) given in the instruction manual.

All the transactions of all the teams are registered by means of a standardized spreadsheet which we provided to the teacher (see Figure A5.1 in the appendix for an example). The transactions include both the financial transactions and other decisions made by the teams, for example with regards to the content of their mission statement, the sustainability trademark and the amount and type of yarn each team buys. On the last day of the program the information from the excel sheet is used to determine the rank of the teams in the class, based on the reward structure applicable (as given by the treatment assignment). To ensure that all teachers calculate and reward the winning teams in each treatment in the same way, the ranking for each team is calculated automatically on a separate sheet of the teacher excel file (see Figure A5.2 in the appendix). A researcher visited each school on the day before the end of the program to bring the gift vouchers for the winning teams and to check their compliance with the course guidelines (including the assigned treatment). Furthermore, we used this opportunity to assist the teachers with the completion of the excel sheet containing all the details of the team results and to encourage response to the second questionnaire, which was to be filled out by the pupils after the program.[12].

3 and 4. For the sake of clarity we describe them again here and highlight the relevant differences for this experiment.

[12]The response rate for both questionnaires was 92,5%

**Table 5.2.** Sample composition

| | Final sample | | |
|---|---|---|---|
| Treatment | # of schools | # of classes | # of teams |
| Baseline | 15 | 15 | 75 |
| non-incentivized CSR | 13 | 18 | 70 |
| incentivized CSR | 16 | 21 | 79 |
| Total | 44 | 54 | 224 |

### 5.3.2 Sample

We conducted the field experiment in the spring of 2010. The participation of schools in the program is voluntary and a school or a class can sign up for it at the beginning of each school year (before January). In January of 2010 the BizWorld foundation provided us with a list of participating schools in the Netherlands. In order to be able to monitor the schools in our experiment closely we decided to focus on schools close to Amsterdam, where our university is located. The total sample includes 46 schools, consisting of 61 classes, that we randomly assigned to the three different treatments. Large schools in the Netherlands sometimes have more than one class in the final grade. In order to avoid spill-over effects between the classes within the same school, the treatment assignment was done at the school level. The final composition of the sample is shown in Table 5.2. This sample includes only those schools and classes for which we received all the required information from the excel sheet completed by the teacher. The overall response rate for the teacher excel files is 88% (54 out of 61 classes).

Some descriptive statistics for the composition of the final sample are shown in the first column of Table 5.3. These descriptives show for example that the sample is equally divided into boys and girls and that the average age is a little under 12. Furthermore, 15% of the children in our sample have (at least) one parent who is not born in the Netherlands. Moreover, we find that the average school test score is 56.26, which is above the nation wide average of 50. In Chapter 3 (Section 3.2.2, Table 3.2) we provide some descriptive evidence on how these results relate to the Dutch population of school kids in the last grade of primary school. The evidence shows that the sample is indeed representative, in terms of individual, school and neighborhood characteristics, for the western part of the Netherlands.

### 5.3.3 Outcome variables

To be able to measure the impact of the different treatments on sustainable behavior and financial performance, we use several outcome variables.

*Sustainable behavior*

As a first indication of differences between the treatments in sustainable behavior we use a dummy variable (*CSR in mission*) to indicate whether or not the team included sustainability (or CSR) into their mission statement. The second sustainability measure, *% sustainable yarn*, is based on the percentage of strands of sustainable yarn bought by the team compared to the total number of strands. For example, if a team buys three strands of yarn, one of which is sustainable, the value of this measure is 33. Finally, we create a dummy variable (*no packaging*) indicating if the team did not buy the plastic packaging material. In Table 5.3 we also provide some descriptive statistics about the use of the sustainability trade mark. However, we do not use it a separate outcome variable in our estimations because it consists of a combination of the two other outcome variables (% sustainable yarn and no packaging).

*Financial performance*

To compare the financial performance of the teams in the different treatments we use *profit* and *revenue*. The revenue is defined as the total earnings from selling the friendship bracelets during the sales market. Profit is measured by calculating the difference between total revenue and the total costs for each team. Not all teams make enough revenue from the sales of their products to cover their costs. Hence for some teams, this number can be negative.

Some descriptive statistics on these outcome variables for the entire sample are shown in the first column of Panel C in Table 5.3.

### 5.3.4 Internal validity

There are several criteria that need to be met to ensure the internal validity of our results. First of all, the assignment to the different treatments has to be independent of potential outcomes. This randomness ensures that the schools in the different treatments are, on average, equal in observed and unobserved characteristics prior to the start of the experiment. To confirm that the treatment assignment was indeed random, we regress several background characteristics on the treatment dummies. For the team characteristics we perform this analysis at the team level, and for the school characteristics we perform the analysis at the school level. The descriptives statistics by treatment are shown in Table 5.3. These results indicate that the randomization was indeed successful. There are only two significant differences between the treatments in terms of background characteristics: female and average team ability. However, the difference for the share of females is only significant between the incentivized CSR treatment and the baseline treatment and not between the two CSR treatments. The differences between the treatments in average ability only seems to be present at the

**Table 5.3.** Descriptives by treatment

| | Total sample | | Baseline | | Non-incentivized CSR | | Incentivized CSR | |
|---|---|---|---|---|---|---|---|---|
| | mean | std. dev. | mean | std. dev. | mean | std. dev. | mean | std. dev. |
| ***A: Background characteristics (team)*** | | | | | | | | |
| % Female | 0.50 | 0.15 | 0.47 | 0.16 | 0.49 | 0.14 | 0.55*** | 0.14 |
| Age | 11.71 | 0.36 | 11.69 | 0.32 | 11.70 | 0.46 | 11.72 | 0.30 |
| Average team ability | 3.01 | 0.71 | 2.82 | 0.60 | 3.31** | 0.65 | 2.92 | 0.78 |
| Team size | 5.79 | 0.58 | 5.79 | 0.55 | 5.74 | 0.56 | 5.84 | 0.63 |
| Nationality parents: One non-dutch | 0.15 | 0.16 | 0.13 | 0.15 | 0.16 | 0.18 | 0.15 | 0.16 |
| Education mother | 2.96 | 0.61 | 2.95 | 0.62 | 2.92 | 0.60 | 2.99 | 0.61 |
| Education father | 3.07 | 0.59 | 3.07 | 0.64 | 3.10 | 0.51 | 3.04 | 0.61 |
| ***B: Background characteristics (school)*** | | | | | | | | |
| Average income per year in € | 20550 | 3990 | 20643 | 4485 | 21230 | 4341 | 19859 | 2977 |
| Average school ability | 56.26 | 9.88 | 53.47 | 7.63 | 59.67 | 11.36 | 55.47 | 9.25 |
| Class size | 22.80 | 5.17 | 23.10 | 6.86 | 24.06 | 4.28 | 21.54 | 4.08 |
| ***C: Outcome variables*** | | | | | | | | |
| CSR in mission (n=173) | 0.46 | 0.50 | 0.19 | 0.40 | 0.41 | 0.50 | 0.67*** | 0.47 |
| Percentage of sustainable yarn purchased | 8.56 | 12.80 | 3.89 | 8.36 | 4.22 | 8.77 | 15.24*** | 15.04 |
| No packaging material purchased | 0.20 | 0.40 | 0.11 | 0.31 | 0.09 | 0.28 | 0.38*** | 0.49 |
| Sustainability trademark (n=165) | 0.21 | 0.41 | 0.13 | 0.34 | 0.07 | 0.25 | 0.36* | 0.48 |
| Profit | 24.77 | 29.87 | 20.81 | 27.51 | 24.11 | 25.71 | 29.12 | 34.83 |
| Revenue | 80.67 | 30.62 | 75.68 | 27.15 | 81.23 | 26.80 | 84.91 | 36.11 |
| Number of teams (schools) | 222 (40) | | 75 (13) | | 70 (12) | | 79 (15) | |

*Note: */**/*** indicates a significant difference at the 10%/5%/1%-level in the team level average of the variable compared to the baseline treatment. Average team ability is proxied by the intended future high school track and ranges from 1 (pre-vocational secondary education) to 5 (pre-university education).Average school ability is calculated by the unweighted average of the test scores on mathematical and verbal ability from a nation wide exam (n=153). Education of the parents is measured in four categories and ranges from 1 (no high school) to 4 (university degree). Observations are clustered at the school level.

team level and not at the school level.[13]

Next, we perform several tests to check for non random attrition from the sample. Attrition is non random if missing results are systematically different between treatments. To this end we perform a regression in which the dependent variable is an indicator for whether or not the team results are observed, and the explanatory variables are the treatment dummies. This regression indicates that there is a marginally significant difference between the missings for the baseline treatment compared to the non-incentivized CSR treatment (p-value: 0.08), with more missing results in the baseline treatment. There are no differences in attrition between the two CSR treatments, and between the incentivized CSR treatment and the baseline treatment.

Furthermore, we use the same method to test for non-random attrition for certain outcome variables of interest. First we look at the dummy variable *CSR in mission*. The results show that this information is (significantly) more often missing for the baseline treatment compared to the incentivized CSR treatment. The difference between the baseline treatment and the non-incentivized CSR treatment is marginally significant (p-value: 0.07), and there are no significant differences in missing mission content between the two CSR treatments. The number of missing observations for the outcome variable *% sustainable yarn* is only significantly different between the baseline treatment and the incentivized CSR treatment and not between the other treatments. Attrition in the CSR outcome variables seems to be negatively correlated with the amount of emphasis placed on it in the treatments. In our case, non random attrition is mainly problematic if it leads to an overestimation of the treatment effect, i.e. if the upward bias in the estimates is so large that we see an effect where there is actually none. This only occurs if the teachers who filled out the complete spreadsheet systematically forgot to fill out the part about the sustainable yarn and the CSR mission content even though the teams in their class used and/or mentioned it. This does not seem to be very likely. As shown in the example in the appendix the requested information is very clear. Had they forgotten about it, they probably would have recalled and checked it when completing the excel sheet. Hence, we are confident that the estimated treatment effects presented below are unbiased (or, if anything, more likely to be an underestimation than an overestimation).

Moreover, to estimate unbiased treatment effect we need to avoid spill-over effects. That is, we have to make sure that the participants from one treatment are not affected by the treatment of another group. To avoid these problems we did the treatment assignment at the school level instead of at the class level. Another important condition

[13]Since the treatment assignment was done at the school level, we have also performed the randomization checks for the team characteristics at the school level. The results are qualitatively the same, but the differences for female and average ability are slightly less significant (p-values: 0.01 and 0.06, respectively).

required for internal validity is that there should be no systematic difference between the treatments across schools. For example, if certain schools pay more attention to a certain part of the program or to a certain part of the treatment (for example by focusing more on sustainability), the treatment is no longer stable across the teams within the same treatment. Another threat to randomized experiments is the substitution bias. This occurs if people who are randomized out of an experiment seek an alternative for the treatment (e.g. if the schools assigned to the baseline treatment would arrange some reward for their teams to stimulate sustainability). However, we are confident that this is not a problem in our experiment. We could not be present during the entire program in all the schools in our sample to monitor their compliance with the treatment directly. However, we visited all the schools and confirmed that the teacher and the entrepreneur followed the guidelines provided in the handbook closely and that there was no deviation from the treatment assignment.

Finally, any experiment could suffer from the Hawthorne effect. That is, if people behave differently because something is different not as a response to the treatment per se. Since all the schools in our sample participate in the same program and are exposed to the same questionnaires, the novel (different than usual) environment is the same for all the teams in the sample. Moreover all classes participate only once (this year) in the program. Hence, any (potential) Hawthorne effect should be eliminated when estimating the treatment effects between the treatments within the same setting. Moreover, the teams and even the teachers in our sample are unaware of the specifics of the design and purpose of the experiment. Therefore, we feel that it is safe to assume that the Hawthorne effect is not a cause of concern in our sample. Taken together these checks ensure that we will be able to estimate the unbiased treatment effect

## 5.4 Results

### 5.4.1 Empirical strategy

In this paper we want to analyze the effect of incentives on sustainable behavior. In particular we are interested in the following question: Do incentives increase sustainable behavior?

To answer this research question we will estimate the following model using a simple OLS estimation.

$$y_i = \beta_0 + \beta_1 incentivizedCSR_i + \beta_2 non\_incentivizedCSR_i + \epsilon_i \tag{5.1}$$

where $y_i$ is the value of the outcome measure of team $i$, $(non\_)incentivizedCSR_i$ is a dummy variable taking the value one if team $i$ was part of the (non-)incentivized CSR

treatment, and zero otherwise and $\epsilon_i$ is an error term. Hence, the baseline treatment serves as our benchmark. Since the observations per school are potentially correlated, we will cluster the observations at the school level in all estimations, to obtain robust standard errors. The treatment assignment was random hence, in theory, our coefficients should not alter when including additional control variables. To confirm the robustness of the results and to increase the precision of the estimated treatment effect, we will also estimate the model with a vector of control variables ($X_i$). We will include several team and school characteristics in our estimation model that might influence the outcome variables. The team characteristics are: average age, gender, ethnic composition, team size, a proxy for average team ability, and the highest education level attained by the parents.[14] Additionally, there are some school characteristics that will be used as explanatory variables that are potentially associated with our outcome variables. These characteristics are class size and the gross average income per year in the neighborhood of the school, based on the four digit postal code.

With this analysis we want to see if discussing sustainability and encouraging to include it in the mission statement (with and without incentives) results in a different, more sustainable, behavior. Secondly, we estimate the (pure) incentive effect on sustainable behavior by comparing the sustainable behavior between the non-incentivized and the incentivized CSR treatment. To this end we perform a Wald test to compare the estimated coefficients between the two CSR treatments (i.e., we test if $\beta_1 = \beta_2$).

Additionally, we test if a trade-off exists between sustainable behavior and financial performance within our experimental setting. To this end we estimate the effect of the different treatments on financial performance and perform the same analysis as described above, but now using the financial performance measures as our outcome variables. The descriptive results of the outcome variables in the different treatments are shown in Figure 5.1 (and in Panel C of Table 5.3 already discussed above). The figures on top show the average of the outcome measures on sustainable behavior per treatment. The average results for financial performance are shown below.

### 5.4.2 Main results

The main estimation results are shown in Table 5.4. The left hand side of the table (columns 1 to 6) show the results for the different measures of sustainable behavior. The results on financial performance can be found on the right hand side (columns 7 to 10). The main results show the comparison between the baseline treatment and the two CSR treatments.

[14] Average ability is measured by the average (self-reported) future high school track of the team members. The high school tracks in the Netherlands range from pre-vocational secondary education to pre-university education.

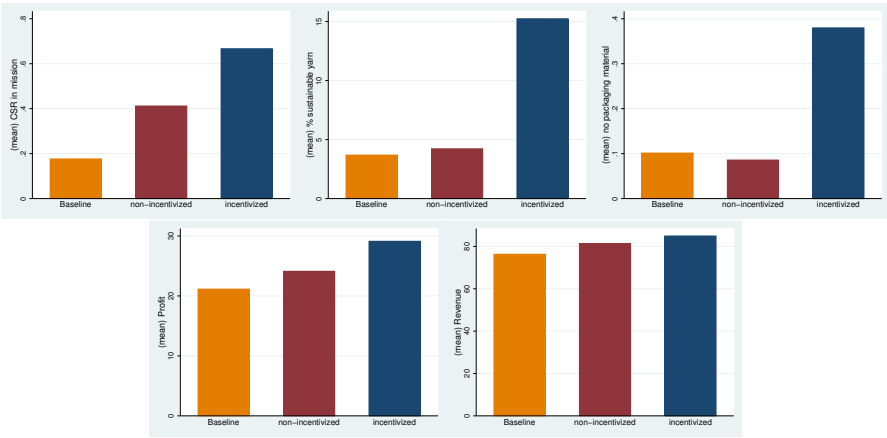
**Table 5.4.** Main results

| | **Sustainable behavior** | | | | | | **Financial performance** | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CSR in mission | | % sustainable yarn | | no packaging | | profit | | revenue | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| incentivized CSR | 0.490*** | 0.401*** | 11.553*** | 12.901*** | 0.278*** | 0.247*** | 7.979 | -1.645 | 8.604 | -1.041 |
| | (0.116) | (0.111) | (2.912) | (3.046) | (0.079) | (0.073) | (8.402) | (5.135) | (8.477) | (4.828) |
| non-incentivized CSR | 0.235 | 0.227 | 0.536 | 0.403 | -0.016 | -0.007 | 2.973 | -1.204 | 4.921 | 0.435 |
| | (0.148) | (0.149) | (2.097) | (2.576) | (0.061) | (0.063) | (5.315) | (5.655) | (5.481) | (5.474) |
| *(omitted category: Baseline)* | | | | | | | | | | |
| Age | | 0.221* | | 4.094 | | -0.031 | | 11.667** | | 14.391*** |
| | | (0.130) | | (3.135) | | (0.087) | | (4.976) | | (4.735) |
| Female | | 0.562*** | | 2.893 | | -0.108 | | 3.247 | | 1.390 |
| | | (0.193) | | (5.266) | | (0.185) | | (12.925) | | (12.882) |
| Average ability | | -0.045 | | 2.955* | | -0.021 | | 6.071** | | 7.282** |
| | | (0.070) | | (1.668) | | (0.043) | | (2.828) | | (2.961) |
| Team size | | 0.028 | | 0.508 | | 0.023 | | 9.616*** | | 14.616*** |
| | | (0.073) | | (2.253) | | (0.050) | | (3.469) | | (3.324) |
| % one parent not NL | | 0.317 | | -1.974 | | -0.127 | | -26.355** | | -25.356** |
| | | (0.244) | | (4.936) | | (0.167) | | (11.071) | | (11.697) |
| Education level mother | | 0.076 | | 1.397 | | -0.036 | | 4.091 | | 3.618 |
| *(1: no high - 4: uni)* | | (0.076) | | (1.913) | | (0.057) | | (3.231) | | (3.056) |
| Education level father | | -0.097 | | -1.243 | | -0.031 | | -2.785 | | -2.468 |
| *(1: no high - 4: uni)* | | (0.079) | | (1.493) | | (0.044) | | (3.973) | | (3.969) |
| Average income per year | | 0.000 | | -0.001* | | 0.000 | | -0.000 | | -0.000 |
| *(based on 4-digit postal code)* | | (0.000) | | (0.000) | | (0.000) | | (0.001) | | (0.001) |
| Class size | | -0.009 | | 0.153 | | -0.005 | | -1.095** | | -1.124** |
| | | (0.012) | | (0.243) | | (0.006) | | (0.508) | | (0.486) |
| Constant | 0.176* | -2.524 | 3.684** | -45.835 | 0.101* | 0.555 | 21.141*** | -161.349** | 76.308*** | -168.144** |
| | (0.089) | (1.852) | (1.382) | (44.247) | (0.055) | (1.141) | (4.142) | (70.604) | (4.098) | (67.590) |
| Number of teams | 177 | 161 | 192 | 175 | 228 | 208 | 227 | 208 | 227 | 208 |
| $R^2$ | 0.169 | 0.245 | 0.189 | 0.257 | 0.119 | 0.115 | 0.013 | 0.127 | 0.014 | 0.181 |
| incent. CSR = non-incent. CSR[1] | | 0.196 | | < 0.001*** | | < 0.001*** | | 0.914 | | 0.730 |

1) This row reports the p-value of Wald tests on $\beta$(incentivized CSR) = $\beta$(non-incentivized CSR).

Standard errors in parentheses. * $p < 0.10$ , ** $p < 0.05$ , *** $p < 0.01$

**Figure 5.1.** Descriptive results by reward treatment




The results show that there are no significant differences in sustainable behavior between the baseline treatment and the non-incentivized CSR treatment. The share of teams that include CSR into their mission statement increases from 19% in the baseline treatment to 41% in the non-incentivized CSR treatment. However, this difference is not significantly different from zero. There is also no significant difference in the % sustainable yarn used by the teams and the teams in both treatments are equally likely to use the plastic packaging material. Apparently, merely discussing sustainability and encouraging the teams to include it into their mission statement does not change the behavior of the children. The comparison between the baseline and the incentivized CSR treatment reveals that this treatment has a significant positive effect on all three sustainable outcome measures. Furthermore, to separate the effect of encouragement on sustainable behavior from the (pure) incentive effect, we also compare the estimated coefficients from the two CSR treatments (non-incentivized vs. incentivized). The p-values of the Wald tests are shown in the bottom row of Table 5.4.

We find that the estimated coefficients for the incentivized CSR treatment are larger (more positive) for all three outcome variables. The treatment effect is significant for % sustainable yarn and the use of no packaging material. These findings indicate that incentives are an effective tool to induce sustainable behavior.

Within the context of our design we also study how sustainable behavior affects financial performance. To this end we compare the financial performance of the teams across the three groups. The results are presented on the right hand side of Table 5.4 and show that there are no significant differences in financial performance between any of the treatments. This indicates that, in this setting, there is no downside to the provision of incentives for CSR, i.e., the increase in sustainable behavior does not lead

**Table 5.5.** Reasons to produce sustainable products

| | Total | Baseline | Non-incentivized CSR | Incentivized CSR |
|---|---|---|---|---|
| Pretty | 0.10 | 0.19 | 0.10 | 0.07 |
| Environment | 0.56 | 0.41 | 0.54 | 0.63 |
| Marketing | 0.16 | 0.21 | 0.24 | 0.11 |
| Team pressure | 0.07 | 0.08 | 0.07 | 0.06 |
| Other | 0.11 | 0.11 | 0.05 | 0.14 |
| Number of observations | 441 | 91 | 96 | 254 |

*Note: Responses to this question are collected at the individual level and are conditional on actual sustainable behavior.

to a lower financial performance.

### 5.4.3 Mechanisms

To find out a little bit more about what drives the children's sustainable behavior, we use a question from the post-test questionnaire. The main question that we use is: "If your team decided to produce sustainable products, why did you do that?". The possible answers were: We thought it was prettier, We found the environment important, To sell more products/for marketing purposes, My team mates wanted it, or Other, please specify. When we look at the differences in drivers for sustainable behavior we see that there are not many differences between the treatments (see Table 5.5).[15] However, there is one noteworthy exception: the environment. In both CSR treatments the children list the environment (significantly) more often as their main motivator to produce sustainable products than in the baseline treatment. This outcome variable was measured at the individual level several weeks after the program. It indicates that the children in these two treatments have a different attitude towards the environment than the children in the baseline treatment. If this learning element has a lasting impact, early attention to the environment with or without incentives to induce sustainable behavior could produce positive spillovers in the future (e.g. Cunha and Heckman, 2007).

### 5.4.4 Robustness checks

We perform several checks to confirm the robustness of the results presented above. First, because the randomization was done at the school level, one could argue that the school rather than the team should be the unit of analysis. We perform the same

[15]The answer to this question is only available for a subsample. Part of the sample in the project described in this paper served as a control group in the evaluation study described in Chapter 3 (see Table A2.1 for an overview of the different treatment assignments). Due to the fact that we used a wait-listed control group approach, we asked those children to fill out the post-test questionnaire before the start of the education program.

estimations at the school level and find that the results are very similar (see Table A5.1 in the Appendix).

Furthermore, some schools did not announce the content of the prize for the winning team or changed it to a prize for the entire class. As this potentially reduces the incentives to win, we include dummy variables to control for potential differences in the incentive intensity of these teams. The estimated coefficients for these dummy variables are not significant in any of the specification. Moreover, the estimated treatment effects do not change when these additional controls are included.

Finally, since this experiment was conducted parallel to the experiment described in Chapter 4, we want to make sure that the treatment variations from these two papers are uncorrelated. To this end, we include a dummy variable for each of these treatment variations. The results show that some of the coefficients on the team treatment dummies are significant. However, the main results remain the same when controlling for these treatments.[16]

## 5.5 Concluding discussion

In this study we test if incentives are an effective tool to induce sustainable behavior in a productive (team) environment in which both financial performance and sustainability matter. To this end, we estimate the effect of three different treatments (baseline, non-incentivized CSR and incentivized CSR) on three sustainable outcome measures. Sustainable behavior is measured in terms of the percentage of sustainable yarn, and two dummy variables: one indicating whether or not a team included CSR into their mission statement and one for the use of plastic packaging material. We also look at the association between sustainable behavior and financial performance (in terms of profit and revenue). To be able to identify the (pure) incentive effect we split our research question into two parts. First, we test if emphasizing and encouraging CSR affects sustainable behavior. Secondly, we provide (equally strong) incentives for financial performance and sustainable behavior and test how this influences sustainable decision making.

The results indicate that there is no difference in terms of sustainable behavior between the teams in the baseline treatment and the teams in the non-incentivized CSR treatment. The estimates from the comparison between the two CSR treatments (i.e., with and without incentives) show that incentives positively affect sustainable behavior. Because a trade-off is often assumed between social and financial performance, we compare the financial results of the teams across the three treatments. The results

[16] A correlation matrix of the different treatment variations is shown in the appendix of Chapter 2 in Table A2.2.

show that there are no significant differences in terms of profit or revenue between the three treatments. This implies that, in this setting, the provision of incentives to induce sustainable behavior does not affect financial performance.

To interpret these results and to judge their external validity, we need to take a closer look at our specific experimental setting. The random treatment assignment is conducted at the school level. Hence, within each school the same reward structure applies. As such, this experimental design can probably best be compared to a situation in which the regulation with respect to CSR within an entire industry changes and not to the choice of an individual company to adopt new rules and/or to introduce incentives. Our baseline treatment then reflects the setting without CSR regulations. The non-incentivized CSR treatment can be compared to a situation with self-regulation. Thus, to a setting in which companies within an industry are encouraged to carefully think about socially responsible behavior (e.g. in terms of the environment, the diversity of their workforce, etc.). Our results indicate that, if there are no clear incentives to actually stimulate corporate social responsibility, encouragement (or self-regulation) alone will not lead to a change in behavior. The incentivized CSR treatment corresponds to a setting where the regulation is enforced and companies are punished if they do not comply, i.e. if they do not behave sustainable. That is, the incentives in this treatment group are designed in such a way that those teams who do not behave sustainable automatically reduce their expected reward, i.e., their chances of winning. The results from this treatment indicate that if suitable performance measures are available, i.e., if CSR can be enforced, the use of these measures can positively affect sustainable behavior. Given the experimental design it is not clear what to expect in terms of financial performance of the individual firms within an industry. That is, if all firms start to produce sustainable products as a result of the regulatory changes, it is not obvious if and how sustainable behavior of an individual firm should affect financial performance. In our experiment, we do not find any differences in financial performance between the treatments.

Obviously, the education program provides us with a very stylized experimental setting that limits the external validity of the results. In section 5.3.2 we provide some descriptive evidence about representativeness of the sample compared to the population. These descriptive statistics indicate that the results presented below are (probably) generalizable for the population of school children aged 11 or 12 in the western part of the Netherlands. However, given the importance of the topic it would be valuable if we could extend our results beyond the children's population. External validity requires that the behavior we observe in our sample of children in response to the different treatments also holds for adults. There is indeed some evidence about the stability in economic behavior from children to adults (Harbaugh et al., 2001). In

their study, Harbaugh et al. (2001) show that at the age of 11 about 60 percent of the children make utility maximizing choices. Furthermore, the results presented in this paper are in line with the results found in several recent field experiments (see Ashraf et al., 2014; Miller et al., 2012; Olken et al., 2012). These studies find that (both financial and non-financial) incentives increase pro-social behavior for adults in various different settings. Moreover, similar to our findings for the comparison between the incentivized and the non-incentivized CSR treatments, Miller et al. (2012) and Olken et al. (2012) also show that only providing information about (how to realize) social outcomes does not lead to an improvement in the level of these outcomes. Finally, in the current experimental setting we only measure the children's direct behavioral responses to the different treatments. However, as a by-product the attitude of the children towards the environment also seems to change. The children in the two CSR treatments list the environment (significantly) more often as their main motivator to produce sustainable products than the children in the baseline treatment. This outcome variable was measured 3 to 8 weeks after the program. If this learning element has a lasting impact, theoretical and empirical research suggests that early attention to the environment (e.g. in primary school) could provide positive spillovers in the future (e.g. Cunha and Heckman, 2007).

# Appendix

Table A5.1: Results school level analysis

| | **Sustainable behavior** | | | | | | **Financial performance** | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CSR in mission | | % sustainable yarn | | no packaging | | profit | | revenue | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| incentivized CSR | 0.448*** | 0.405*** | 11.466*** | 13.474*** | 0.231** | 0.305*** | 4.576 | 0.023 | 4.742 | -0.321 |
| | (0.130) | (0.131) | (3.529) | (4.264) | (0.095) | (0.107) | (8.065) | (7.171) | (7.925) | (6.515) |
| non-incentivized CSR | 0.255 | 0.311* | 1.232 | 0.913 | -0.037 | 0.005 | 0.388 | -0.737 | 1.545 | -0.280 |
| | (0.157) | (0.152) | (2.414) | (2.736) | (0.075) | (0.105) | (6.382) | (7.486) | (6.278) | (7.349) |
| *(omitted category: Baseline)* | | | | | | | | | | |
| Age | | 0.128 | | 6.845 | | 0.002 | | 2.436 | | 4.595 |
| | | (0.187) | | (5.051) | | (0.104) | | (6.317) | | (6.815) |
| Female | | 0.726 | | -14.425 | | -1.076** | | -20.703 | | -15.948 |
| | | (0.715) | | (20.597) | | (0.416) | | (31.642) | | (28.602) |
| Average ability | | -0.062 | | 6.740 | | -0.052 | | 8.625 | | 10.195 |
| | | (0.200) | | (5.003) | | (0.095) | | (6.613) | | (6.899) |
| Team size | | 0.089 | | 4.680 | | -0.015 | | 7.267 | | 13.335** |
| | | (0.155) | | (4.459) | | (0.096) | | (6.208) | | (6.189) |
| % one parent not NL | | 1.294* | | 29.989** | | 0.294 | | -41.190 | | -36.035 |
| | | (0.722) | | (14.570) | | (0.366) | | (25.747) | | (25.672) |
| Education level mother | | 0.319 | | 2.516 | | -0.030 | | -5.258 | | -6.580 |
| *(1: no high - 4: uni)* | | (0.203) | | (3.697) | | (0.165) | | (9.360) | | (8.851) |
| Education level father | | -0.113 | | -4.988 | | -0.038 | | -10.619 | | -9.123 |
| *(1: no high - 4: uni)* | | (0.243) | | (3.832) | | (0.131) | | (8.747) | | (9.041) |
| Average income per year | | -0.000 | | -0.001 | | 0.000* | | 0.000 | | 0.000 |
| *(based on 4-digit postal code)* | | (0.000) | | (0.001) | | (0.000) | | (0.001) | | (0.001) |
| Class size | | -0.004 | | 0.112 | | 0.003 | | -1.174** | | -1.226** |
| | | (0.015) | | (0.270) | | (0.007) | | (0.575) | | (0.533) |
| Constant | 0.205** | -2.460 | 3.931** | -98.028 | 0.133* | 0.477 | 24.290*** | 12.373 | 79.794*** | 3.240 |
| | (0.094) | (2.646) | (1.612) | (72.608) | (0.068) | (1.459) | (5.476) | (88.768) | (5.131) | (95.456) |
| Number of teams | 38 | 37 | 39 | 38 | 45 | 44 | 45 | 44 | 45 | 44 |
| $R^2$ | 0.226 | 0.412 | 0.270 | 0.526 | 0.214 | 0.343 | 0.011 | 0.314 | 0.011 | 0.342 |
| incent. CSR = non-incent. CSR[1] | | 0.541 | | 0.004*** | | < 0.001*** | | 0.892 | | 0.994 |

1) This row reports the p-value of Wald tests on $\beta$(incentivized CSR) = $\beta$(non-incentivized CSR).
Standard errors in parentheses. * $p < 0.10$ , ** $p < 0.05$ , *** $p < 0.01$

Figure A5.1: 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 spent** | | | **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 | # own shares | **Team company value** | number of team members |
| Friends & Co. | 269 1 | 104 | 7 | 1 | 10 | 86 | 97 | 60 | 37 | 11 | 26 | 7 | 182 | 6 |
| V and B | 269 2 | 59 | 7 | 1 | 10 | 41 | 55 | 61 | -6 | 0 | 0 | 6 | 0 | 6 |
| Made by Uzz! | 269 3 | 84 | 7 | 1 | 10 | 66 | 82 | 63 | 19 | 5 | 14 | 7 | 98 | 6 |

Figure A5.2: Example calculation final score (incentivized CSR treatment)

**Sustainability points**

| Company/Team name: | Team number | points for sustainable yarn | points for not using packaging material | CSR bonus points (max. 50 points per team) | **Total sustainability points** | Team rank based on sustainability |
|---|---|---|---|---|---|---|
| Different St | 228 1 | 10 | 0 | 0 | 10 | 5 |
| B.B. | 228 2 | 12 | 0 | 0 | 12 | 3 |
| Extreme St | 228 3 | 21 | 60 | 40 | 121 | 1 |
| Respect | 228 4 | 27 | 60 | 30 | 117 | 2 |
| Bizzzy | 228 5 | 11 | 0 | 0 | 11 | 4 |

**Final score (profit and sustainability)**

| Company/Team name | Average ranking (profit rank + sustainability rank)/2 | **Team rank based on profit and sustainability** |
|---|---|---|
| Different St | 3 | 2 |
| B.B. | 3,5 | 4 |
| Extreme St | 2 | 1 |
| Respect | 3,5 | 4 |
| Bizzzy | 3 | 2 |

**Team company value**

| Company/Team name: | Team number | Profit/Loss | Company value (Profit or Loss x 10) | Price per share (Company value/total # of shares) | # of shares owned by the team | **Team company value** | Team rank based on profit |
|---|---|---|---|---|---|---|---|
| Different St | 228 1 | 38 | 380 | 38 | 7 | 266 | 1 |
| B.B. | 228 2 | 10 | 100 | 10 | 7 | 70 | 4 |
| Extreme St | 228 3 | 12 | 120 | 12 | 7 | 84 | 3 |
| Respect | 228 4 | -28 | -280 | -28 | 7 | -196 | 5 |
| Bizzzy | 228 5 | 14 | 140 | 14 | 7 | 98 | 2 |

# Chapter 6

# Concluding remarks

This dissertation reports the results of three field experiments. Chapter 3 analyzes the effectiveness of early entrepreneurship education. Given the key role that entrepreneurial activity has in fostering economic growth and innovation, the evaluation of measures that may stimulate successful entrepreneurship is of interest to both academics and practitioners. Chapter 3 discusses such an evaluation. It reports the results from a study that evaluates a leading entrepreneurship education program that is taught worldwide in the final grade of primary school. It is the first evaluation study of an entrepreneurship education program for children in primary school. Other studies evaluated entrepreneurship education programs mostly in tertiary and sometimes in secondary education. The findings indicate that the program studied in this dissertation has a robust positive effect on non-cognitive entrepreneurial skills. This result is surprising since previous evaluations found zero or negative effects (for older pupils or students). The insignificant effects found in previous studies may well be due to the fact that entrepreneurial skills and knowledge are more easily developed earlier in life or because the returns to training programs later in life depend on investments in knowledge and skills made earlier. Thus, the results presented in this chapter tentatively suggest that it might be more effective to develop non-cognitive entrepreneurial skills at an early age. One of the follow-up questions that is raised by this study concerns the optimal age and timing to develop non-cognitive entrepreneurial skills. The results obtained provide a relevant first step for future research to investigate whether the life-cycle model of skill formation indeed holds for the development of entrepreneurial skills as well. Another related question that remains unanswered by this study pertains to the long term effects of entrepreneurship education. It is important, both from a policy perspective as well as from a general research interest, to further study and understand the costs and benefits of entrepreneurship education. Nevertheless, finding short term effects is a first step towards a better understanding of the effects of entrepreneurship education, including potential dynamic spillover effects.

Chapter 4 investigates how skill composition affects the performance of teams. Teamwork and team entrepreneurship is a growing phenomenon. Moreover, the tasks within all types of organizations (from start-ups to large firms) have become increasingly complex. Complex environments require the combination of different types of knowledge and skills in order to be successful. The combination of an increase in task complexity and the observed increase in teamwork raises the question of effective team composition. The aim of this chapter is to answer two specific research questions: (1) Do teams consisting of a combination of individual generalists (JATs) perform better than teams consisting of (one type of) individual specialists? and (2) Can a lack of skill balance at the individual level be compensated by combining different types of specialists within one team? The findings presented in this chapter show that teams benefit from having balanced skills, but that it is difficult to substitute a lack of skill balance at the individual level by combining the skills of different specialists within one team. As this result was obtained in a field experiment studying teams of children participating in an entrepreneurship education program, more research on this topic is required to confirm that the results are indeed robust in other settings. There is also scope for future research on entrepreneurial teams through the use of observational data. One interesting question here is what drives the choices of individuals to work or start-up in a team or alone, and how this affects subsequent performance. Another topic that deserves further investigation is how entrepreneurial teams develop over the life cycle of the firm. Different stages in the start-up phase require different skills. It would be interesting to see if the composition of entrepreneurial teams changes accordingly and how this relates to team performance.

Chapter 5 investigates whether incentives are an effective tool to induce sustainable behavior. The topic is motivated by some recent field experiments on social behavior and the more descriptive evidence on the association between incentives and CSR. The experimental design used in this study reflects a productive (team) environment in which both financial performance and sustainability matter. The results from this experiment indicate that incentives have a significant positive effect on sustainable behavior. However, the comparison between the baseline treatment and the non-incentivized CSR treatment suggests that merely discussing the environment and emphasizing its importance, does not cause a behavioral change. Moreover, descriptive evidence indicates that the two CSR treatments (positively) influence the children's attitude towards the environment. More research is needed to see if the results found for children at the age of 11 or 12 indeed translate to adults as well. In future research it would be interesting to conduct field experiments to test the effect of incentives on sustainable behavior in real life business environments where a potential trade-off exists between financial performance and sustainable behavior. Furthermore, similar to the

results found on the the development of entrepreneurial skills, it would be valuable to see the long term effects of lessons on sustainability in primary school on sustainable behavior and environmental attitudes later in life.

The implications of the results presented in this dissertation strongly depend on the internal and the external validity of these results. Field experiments provide unique settings that allow the estimation of causal effects for research questions that are otherwise troubled with methodological difficulties. However, the high internal validity of this method (potentially) comes at a cost of lower external validity.

The trade-off between internal and external validity is most pronounced for the topics presented in Chapter 4 and Chapter 5. If one were allowed to pick an ideal setting to study team composition and sustainable behavior, an entrepreneurship education program in primary school is probably not the first setting that comes to mind. However, the program does provide an attractive, controlled environment where these two topics can be studied in a stylized setting. Advantages of this particular setting to study the effect of balanced skills on team performance, as is done in Chapter 4, include the measurability of the relevant skill set and the possibility to compose the teams exogenously (which allows us to study an interesting causal effect). Moreover, (a priori) there is no clear reason to assume that the treatment effects found for children should vary substantially with age or subject pool. Another possible limitation for the results presented in Chapter 4 is that we define JATs and specialists only in terms of two types of skills: mathematical and verbal ability. We argue that since these two skills are considered the two core skills for children at the age of 12, they form the basis of relevant skills to be developed later in life.

As argued in Chapter 5, the results on sustainable behavior presented there are (probably) generalizable for the population of school children aged 11 or 12 in the western part of the Netherlands. However, given the importance of the topic it would be valuable if we could extend our results beyond the children's population. External validity requires that the behavior we observe in our sample of children in response to the different treatments also holds for adults. There is indeed some evidence for the stability in economic behavior from children to adults; Harbaugh et al. (2001) show that at the age of 11 about 60 percent of the children make utility maximizing choices. Furthermore, the results presented in Chapter 5 are in line with the results found in several recent field experiments (e.g. Ashraf et al., 2014). These studies also find that incentives (both financial and non-financial) increase pro-social behavior. Nevertheless, some caution is required not to draw too strong inferences from our results for the behavior of adults.

For the results presented in Chapter 3 and for those presented in Chapter 5 on the environmental behavior and attitude of *children*, the external validity relies on two

assumptions: (1) the program in the sample is a typical example of the education program under investigation, and (2) the sample itself is representative for the population studied. We conduct several (descriptive) tests to confirm that these two assumptions indeed hold and we find no indication that either of them is violated. Hence, the results presented in Chapter 3 imply that primary school might be a fruitful environment to teach children non-cognitive skills. Moreover, recent economic research indicates that the non-cognitive skills studied here are relevant beyond the entrepreneurship setting, i.e., for labour market outcomes in general (Heckman et al., 2013, 2006). Taken together, these findings suggest that it could be beneficial to extend the current focus on core skills such as mathematical and verbal ability, by incorporating important non-cognitive skills into the primary school curriculum as well.

The findings presented in Chapter 5 suggest that incentives are required to induce sustainable behavior at a young age (11 to 12). Moreover, the two CSR treatments seem to have a positive effect on the individual environmental attitudes of the children in these treatments. These results suggest that even though discussing and encouraging CSR does not lead to a direct behavioral change, it does have a positive influence on the children's attitude towards the environment. If this learning element has a lasting impact, theoretical and empirical research suggests that early attention to the environment (e.g. in primary school) could provide positive spillovers in the future (e.g. Cunha and Heckman, 2007).

# Bibliography


Almlund, M., Duckworth, A., Heckman, J., and Kautz, T. (2011). Personality psychology and economics. Technical report, IZA Discussion paper series.

Ariely, D., Bracha, A., and Meier, S. (2009). Doing good or doing well? image motivation and monetary incentives in behaving prosocially. *The American Economic Review*, pages 544–555.

Ashraf, N., Bandiera, O., and Jack, K. (2014). No margin, no mission?: A field experiment on incentives for pro-social tasks.

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

Bandiera, O., Barankay, I., and Rasul, I. (2011). Field experiments with firms. *The Journal of Economic Perspectives*, pages 63–82.

Barbaranelli, C., Caprara, G., Rabasca, A., and Pastorelli, C. (2003). A questionnaire for measuring the big five in late childhood. *Personality and Individual Differences*, 34(4):645–664.

Baron, R. and Ensley, M. (2006). Opportunity recognition as the detection of meaningful patterns: Evidence from comparisons of novice and experienced entrepreneurs. *Management Science*, 52(9):1331–1344.

Bates, T. (1990). Entrepreneur human capital inputs and small business longevity. *The Review of Economics and Statistics*, pages 551–559.

Baum, J. and Locke, E. (2004). The relationship of entrepreneurial traits, skill, and motivation to subsequent venture growth. *Journal of Applied Psychology*, 89(4):587–598.

Baum, J., Locke, E., and Smith, K. (2001). A multidimensional model of venture growth. *Academy of Management Journal*, 44(2):292–303.

Becker, G. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. University of Chicago Press.

Begley, T. and Boyd, D. (1987). Psychological characteristics associated with performence in entrepreneurial firms and smaller businesses. *Journal of Business Venturing*, 2(1):79–93.

Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *The American Economic Review*, pages 1652–1678.

Bénabou, R. and Tirole, J. (2010). Individual and corporate social responsibility. *Economica*, 77(305):1–19.

Besley, T. and Ghatak, M. (2013). Profit with purpose? a theory of social enterprise with experimental evidence.

Borghans, L., Duckworth, A., Heckman, J., and Ter Weel, B. (2008). The Economics and Psychology of Cognitive and Non-Cognitive Traits. *Journal of Human Resources*, 43(4):972–1059.

Borghans, L., ter Weel, B., and Weinberg, B. (2006). People people: Social capital and the labor-market outcomes of underrepresented groups. Working Paper 11985, National Bureau of Economic Research.

Bradler, C., Dur, R., Neckermann, S., and Non, A. (2013). Employee recognition and performance: A field experiment. *Tinbergen Institute Discussion paper*.

Brockhaus, R. (1980). Risk taking propensity of entrepreneurs. *Academy of Management Journal*, 23(3):509–520.

Caliendo, M., Fossen, F., and Kritikos, A. (2010). The impact of risk attitudes on entrepreneurial survival. *Journal of Economic Behavior & Organization*, 76(1):45–63.

Cantner, U., Goethner, M., and Stuetzer, M. (2010). Disentangling the effects of new venture team functional heterogeneity on new venture performance. *Jena Economic Research Papers*, 029.

Chen, C., Greene, P., and Crick, A. (1998). Does entrepreneurial self-efficacy distinguish entrepreneurs from managers? *Journal of Business Venturing*, 13(4):295 – 316.

Chetty, R., Friedman, J., Hilger, N., Saez, E., Schanzenbach, D., and Yagan, D. (2011). How does your kindergarten classroom affect your earnings? evidence from project star. *The Quarterly Journal of Economics*, 126(4):1593–1660.

Colombier, N. and Masclet, D. (2008). Intergenerational correlation in self employment: Some further evidence from french echp data. *Small Business Economics*, 30(4):423–437.

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.

Coombs, J. E. and Gilley, K. M. (2005). Stakeholder management as a predictor of ceo compensation: Main effects and interactions with financial performance. *Strategic Management Journal*, 26(9):827–840.

Cortina, J. (1993). What is coefficient alpha? an examination of theory and applications. *Journal of Applied Psychology*, 78(1):98.

Cramer, J., Hartog, J., Jonker, N., and Van Praag, C. (2002). Low risk aversion encourages the choice for entrepreneurship: an empirical test of a truism. *Journal of Economic Behavior & Organization*, 48(1):29–36.

Cunha, F. and Heckman, J. (2007). The technology of skill formation. *The American Economic Review*, 97(2):31–47.

Cunha, F. and Heckman, J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43(4):738–782.

Cunha, F., Heckman, J., and Schennach, S. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931.

Dahl, M. and Sorenson, O. (2012). Home sweet home: Entrepreneurs' location choices and the performance of their ventures. *Management Science*, 58(6):1059–1071.

Dahlin, K. B., Weingart, L. R., and Hinds, P. J. (2005). Team diversity and information use. *Academy of Management Journal*, 48(6):1107–1123.

Davidsson, P. and Honig, B. (2003). The role of social and human capital among nascent entrepreneurs. *Journal of Business Venturing*, 18(3):301–331.

Deckop, J., Merriman, K., and Gupta, S. (2006). The effects of ceo pay structure on corporate social performance. *Journal of Management*, 32(3):329–342.

Dunn, T. and Holtz-Eakin, D. (2000). Financial capital, human capital, and the transition to self-employment: Evidence from intergenerational links. *Journal of Labor Economics*, 18(2):282–305.

Fairlie, R., Karlan, D., and Zinman, J. (2012). Behind the gate experiment: Evidence on effects of and rationales for subsidized entrepreneurship training. *NBER Working Paper No. 17804*.

Fairlie, R. and Robb, A. (2007). Families, human capital, and small business: Evidence from the characteristics of business owners survey. *Industrial and Labor Relations Review*, pages 225–245.

Franck, E. and Nüesch, S. (2010). The effect of talent disparity on team productivity in soccer. *Journal of Economic Psychology*, 31(2):218–229.

Frey, B. S. and Oberholzer-Gee, F. (1997). The cost of price incentives: An empirical analysis of motivation crowding-out. *The American economic review*, 87(4):746–755.

Garcia-Castro, R., Ariño, M., and Canela, M. (2010). Does social performance really lead to financial performance? accounting for endogeneity. *Journal of business ethics*, 92(1):107–126.

Glaeser, E., Laibson, D., and Sacerdote, B. (2002). An economic approach to social capital*. *The Economic Journal*, 112(483):F437–F458.

Gneezy, U. and Rustichini, A. (2000a). Fine is a price, a. *J. Legal Stud.*, 29(1):1–18.

Gneezy, U. and Rustichini, A. (2000b). Pay enough or don't pay at all. *Quarterly journal of economics*, pages 791–810.

Gould, E. and Winter, E. (2009). Interactions between workers and the technology of production: Evidence from professional baseball. *The Review of Economics and Statistics*, 91(1):188–200.

Hamilton, B., Nickerson, J., and Owan, H. (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy*, 111(3):465–497.

Hansen, K., Heckman, J., and Mullen, K. (2004). The effect of schooling and ability on achievement test scores. *Journal of Econometrics*, 121:39 – 98.

Harbaugh, W. T., Krause, K., and Berry, T. R. (2001). Garp for kids: On the development of rational choice behavior. *American Economic Review*, 91:1539–1545.

Harrison, G. and List, J. (2004). Field experiments. *Journal of Economic Literature*, 42(4):1009–1055.

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.

Hausman, J. (1978). Specification tests in econometrics. *Econometrica: Journal of the Econometric Society*, pages 1251–1271.

Heckman, J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 312(5782):1900.

Heckman, J., Pinto, R., and Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, forthcoming:1–52.

Heckman, J., Stixrud, J., and Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3):411–482.

Hoever, I. J., Van Knippenberg, D., van Ginkel, W. P., and Barkema, H. G. (2012). Fostering team creativity: Perspective taking as key to unlocking diversity's potential. *Journal of Applied Psychology*, 97(5):982.

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., and Van Praag, C. (2012). Ability dispersion and team performance: A field experiment. *IZA Discussion Paper*.

Hoogendoorn, S. and Van Praag, C. (2012). Ethnic diversity and team performance: a randomized field experiment. *IZA Discussion Paper*.

Hornaday, J. and Aboud, J. (1971). Characteristics of successful entrepreneurs. *Personnel Psychology*, 24(2):141–153.

Hvide, H. and Panos, G. (2013). Risk tolerance and entrepreneurship. *Journal of Financial Economics*, forthcoming.

Imbens, G. and Wooldridge, J. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1):5–86.

Kahane, L., Longley, N., and Simmons, R. (2013). The effects of coworker heterogeneity on firm-level output: assessing the impacts of cultural and language diversity in the national hockey league. *Review of Economics and Statistics*, 95(1):302–314.

Kanbur, S. (1979). Of risk taking and the personal distribution of income. *The Journal of Political Economy*, pages 769–797.

Karlan, D. and Valdivia, M. (2011). Teaching entrepreneurship: Impact of business training on microfinance clients and institutions. *Review of Economics and Statistics*, 93(2):510–527.

Kihlstrom, R. and Laffont, J. (1979). A general equilibrium entrepreneurial theory of firm formation based on risk aversion. *The Journal of Political Economy*, pages 719–748.

Kilduff, M., Angelmar, R., and Mehra, A. (2000). Top management-team diversity and firm performance: Examining the role of cognitions. *Organization Science*, 11(1):21–34.

Kirschner, F., Paas, F., Kirschner, P. A., and Janssen, J. (2011). Differential effects of problem-solving demands on individual and collaborative learning outcomes. *Learning and Instruction*, 21(4):587–599.

Kitzmueller, M. and Shimshack, J. (2012). Economic perspectives on corporate social responsibility. *Journal of Economic Literature*, 50(1):51–84.

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.

Knight, F. (1921). *Risk, Uncertainty, and Profit*. New York: Houghton-Mifflin.

Knudsen, E., Heckman, J., Cameron, J., and Shonkoff, J. (2006). Economic, neurobiological, and behavioral perspectives on building americas future workforce. *Proceedings of the National Academy of Sciences*, 103(27):10155–10162.

Lazear, E. (1999). Globalization and the market for teammates. *The Economic Journal*, 109(454):15–40.

Lazear, E. (2004). Balanced skills and entrepreneurship. *The American Economic Review*, 94(2):208–211.

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

Lazear, E. P. (2000). Performance pay and productivity. *The American Economic Review*, 90(5):1346–1361.

Lazear, E. P. and Shaw, K. L. (2007). Personnel economics: The economist's view of human resources. *The Journal of Economic Perspectives*, pages 91–114.

Lechler, T. (2001). Social interaction: A determinant of entrepreneurial team venture success. *Small Business Economics*, 16(4):263–278.

Lechmann, D. and Schnabel, C. (2011). Are the self-employed really jacks-of-all-trades? testing the assumptions and implications of lazear's theory of entrepreneurship with german data. *IZA Discussion Paper*.

Leonard, J., Levine, D., and Joshi, A. (2004). Do birds of a feather shop together? the effects on performance of employees' similarity with one another and with customers. *Journal of Organizational Behavior*, 25(6):731–754.

Lindquist, M., Sol, J., and Van Praag, C. (2013). Why do entrepreneurial parents have entrepreneurial children? *Journal of Labor Economics*, forthcoming.

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., and d Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of educational research*, 66(4):423–458.

Mahoney, L. and Thorne, L. (2005). Corporate social responsibility and long-term compensation: Evidence from canada. *Journal of Business Ethics*, 57(3):241–253.

Margolis, J. D., Elfenbein, H. A., and Walsh, J. P. (2007). Does it pay to be good? a meta-analysis and redirection of research on the relationship between corporate social and financial performance. *Ann Arbor*, 1001:48109–1234.

McClelland, D. (1965). N achievement and entrepreneurship: A longitudinal study. *Journal of Personality and Social Psychology*, 1(4):389–392.

McCrae, R., Costa, P., Terracciano, A., Parker, W., Mills, C., De Fruyt, F., and Mervielde, I. (2002). Personality trait development from age 12 to age 18: Longitudinal, cross-sectional, and cross-cultural analyses. *Journal of Personality and Social Psychology*, 83(6):1456–1468.

McGuire, J., Dow, S., and Argheyd, K. (2003). Ceo incentives and corporate social performance. *Journal of Business Ethics*, 45(4):341–359.

McKenzie, D. and Woodruff, C. (2014). What are we learning from business training and entrepreneurship evaluations around the developing world? *The World Bank Research Observer*, 29(1):48–82.

McWilliams, A. and Siegel, D. (2000). Corporate social responsibility and financial performance: correlation or misspecification? *Strategic management journal*, 21(5):603–609.

Miller, G., Luo, R., Zhang, L., Sylvia, S., Shi, Y., Foo, P., Zhao, Q., Martorell, R., Medina, A., and Rozelle, S. (2012). Effectiveness of provider incentives for anaemia reduction in rural china: a cluster randomised trial. *BMJ: British Medical Journal*, 345.

Mincer, J. (1958). Investment in human capital and personal income distribution. *The Journal of Political Economy*, 66(4):281–302.

Olken, B. A., Onishi, J., and Wong, S. (2012). Should aid reward performance? evidence from a field experiment on health and education in indonesia. Technical report, National Bureau of Economic Research.

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.

Pai, H.-H., Sears, D., and Maeda, Y. (2014). Effects of small-group learning on transfer: a meta-analysis. *Educational Psychology Review*, pages 1–24.

Parker, S. (2008). Entrepreneurship among married couples in the united states: A simultaneous probit approach. *Labour Economics*, 15(3):459 – 481.

Parker, S. (2009). *The economics of entrepreneurship*. Cambridge University Press.

Peterman, N. and Kennedy, J. (2003). Enterprise education: Influencing students perceptions of entrepreneurship. *Entrepreneurship Theory and Practice*, 28(2):129–144.

Pfeiffer, F. and Reuss, K. (2008). Age-dependent skill formation and returns to education. *Labour Economics*, 15(4):631–646.

Plass, J. L., O'Keefe, P. A., Homer, B. D., Case, J., Hayward, E. O., Stein, M., and Perlin, K. (2013). The impact of individual, competitive, and collaborative mathematics game play on learning, performance, and motivation. *Journal of Educational Psychology*, 105(4):1050–1066.

Revelle, W. and Zinbarg, R. (2009). Coefficients alpha, beta, omega, and the glb: Comments on sijtsma. *Psychometrika*, 74(1):145–154.

Richard, O., Barnett, T., Dwyer, S., and Chadwick, K. (2004). Cultural diversity in management, firm performance, and the moderating role of entrepreneurial orientation dimensions. *Academy of Management Journal*, 47(2):255–266.

Roberts, P. and Sterling, A. (2012). Network progeny? prefounding social ties and the success of new entrants. *Management Science*, 58(7):1292–1304.

Robinson, P. B. and Sexton, E. A. (1994). The effect of education and experience on self-employment success. *Journal of Business Venturing*, 9(2):141–156.

Rulke, D. L. and Galaskiewicz, J. (2000). Distribution of knowledge, group network structure, and group performance. *Management Science*, 46(5):612–625.

Schumpeter, J. (1934). *Capitalism, socialism, and democracy*. Harper and Row.

Sexton, D. and Bowman, N. (1985). The entrepreneur: A capable executive and more. *Journal of Business Venturing*, 1(1):129–140.

Shane, S. (2003). *A general theory of entrepreneurship: The individual-opportunity nexus*. Edward Elgar Publishing.

Shane, S. and Venkataraman, S. (2000). The promise of entrepreneurship as a field of research. *Academy of Management Review*, 25(1):217–226.

Silva, O. (2007). The jack-of-all-trades entrepreneur: Innate talent or acquired skill? *Economics Letters*, 97(2):118–123.

Souitaris, V., Zerbinati, S., and Al-Laham, A. (2007). Do entrepreneurship programmes raise entrepreneurial intention of science and engineering students? The effect of learning, inspiration and resources. *Journal of Business Venturing*, 22(4):566–591.

Stewart, W. and Roth, P. (2001). Risk propensity differences between entrepreneurs and managers: a meta-analytic review. *Journal of Applied Psychology*, 86(1):145.

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 Dijk, H., van Engen, M. L., and van Knippenberg, D. (2012). Defying conventional wisdom: A meta-analytical examination of the differences between demographic and job-related diversity relationships with performance. *Organizational Behavior and Human Decision Processes*, 119(1):38–53.

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.

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

Waddock, S. A. and Graves, S. B. (1997). The corporate social performance-financial performance link. *Strategic management journal*, 18(4):303–319.

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.

Walls, J. L., Berrone, P., and Phan, P. H. (2012). Corporate governance and environmental performance: is there really a link? *Strategic Management Journal*, 33(8):885–913.

Ward, T. (2004). Cognition, creativity, and entrepreneurship. *Journal of Business Venturing*, 19(2):173–188.

Zhao, H., Seibert, S., and Lumpkin, G. (2010). The relationship of personality to entrepreneurial intentions and performance: A meta-analytic review. *Journal of Management*, 36(2):381.

# Summary (in English)

The aim of this dissertation is to answer three research questions: (1) Can entrepreneurship be taught (in school, when young)?, (2) What is the effect of balanced skills on team performance?, and (3) Can incentives induce sustainable behavior?

**Research question 1**

Chapter 3 is based on the paper "The effect of early entrepreneurship education: Evidence from a field experiment", co-authored with Randolph Sloof and Mirjam Van Praag (European Economic Review 72, 2014, 76–97). The aim of this chapter is to analyze the effectiveness of *early* entrepreneurship education. To this end, a leading entrepreneurship education program that is taught worldwide in the final grade of primary school is evaluated.

The question if entrepreneurship can be taught has been the subject of discussion for many years (e.g., Lindquist et al., 2013; Colombier and Masclet, 2008). The sharp increase in the number of entrepreneurship education programs suggests that the general consensus is that entrepreneurship can indeed be taught. From a policy perspective this is an appealing thought. The idea that entrepreneurs are not necessarily born but can also be developed creates a window of opportunity for (educational) policies aimed at enhancing entrepreneurship. However, there is little research on the effectiveness of such educational programs. Chapter 3 of this dissertation evaluates the effectiveness of an early entrepreneurship education program. A theoretical motivation to look at *early* entrepreneurship education is provided by Cunha and Heckman's (2007) general model of the technology of skill formation. This model emphasizes the importance of early investments in both cognitive and non-cognitive skills. It strongly suggests that an investment in skills not only has a direct impact on the current stock of skills, but also produces spill-over effects in subsequent periods by boosting current skills and by making investments later in life more productive. Early investments in skills may thus be particularly effective in the long run. Obviously, the (potential) future spill-over benefits of early investments in skills only occur if the early investment has a direct impact on the stock of skills in the first place. Chapter 3 of this dissertation therefore evaluates the *direct* (short term) effect of early entrepreneurship education.

To assess the impact of the program we focus on pupils' development of entrepreneurship knowledge and a set of non-cognitive skills relevant for entrepreneurial activity. The results indicate that knowledge is unaffected by the program. However, we find that the program has a robust positive effect on seven (out of nine) non-cognitive entrepreneurial skills. Self-reported scores on (constructs of) Risk taking propensity, Creativity, Need for Achievement, Self-Efficacy, Pro-activity, Persistence and Analyzing all increase significantly more in the treatment group than in the control group. This is surprising since previous evaluations found zero or negative effects. Because these earlier studies all pertain to entrepreneurship education for adolescents, our result tentatively suggests that non-cognitive entrepreneurial skills are best developed at an early age.

**Research question 2**

Chapter 4 is based on the paper "Jacks-of-all-trades? The effect of balanced skills on team performance", co-authored with Randolph Sloof and Mirjam Van Praag. The aim of this chapter is to test how skill composition affects team performance and whether (a lack of) individual balanced skills can be substituted by combining the skills of various specialists within one team.

The research question addressed in this chapter is motivated by the observation that nowadays, teamwork is an omnipresent phenomenon within firms of all types and sizes (Hamilton et al., 2003). Large firms increasingly rely on the work and decisions made by (self-managed) teams (Lazear and Shaw, 2007). Moreover, a substantial and growing share of businesses are started up and run by entrepreneurial teams instead of solo entrepreneurs (Klotz et al., 2014; Parker, 2009). Over the past decades the tasks within all types of organizations (from new ventures to established firms) have become increasingly complex due to new technologies and rapidly changing environments (Dahlin et al., 2005; Lazear and Shaw, 2007). The combination of an increase in task complexity and the observed increase in teamwork raises the question of effective team composition. This question, focusing on the *skill* composition of successful (entrepreneurial) teams, is empirically explored in Chapter 4 of this dissertation.

This chapter is partly motivated by the findings presented by Rulke and Galaskiewicz (2000). They find that the teams of generalists outperform the specialist teams. We add to their study by (among other things) distinguishing between two specific *types* of specialists (math and verbal specialists). Based on these types we explicitly compose teams consisting of different combinations of specialists and compare their performance to the generalist teams. We conduct a field experiment to analyze the role and substitutability of balanced skills for teams within a dynamic (entrepreneurial) environment. To this end, we study teams of children who set up a toy business in

friendship bracelets in an entrepreneurship education program ("BizWorld") in the last grade of primary school in the Netherlands. Based on pupils' precisely measured level of verbal and mathematical ability, we exogenously compose 179 teams separated into four different team types: JAT teams, math-specialist teams, verbal-specialist teams and mixed-specialist teams. Our results show that balanced skills are beneficial to team performance, and that it is hard to substitute individual balanced skills by combining different specialists within one team.

**Research question 3**

Chapter 5 is based on the paper "The effect of incentives on sustainable behavior: Evidence from a field experiment", also co-authored with Randolph Sloof and Mirjam Van Praag. This chapter investigates how to induce sustainable behavior in a productive setting and the link between sustainable behavior and financial performance.

The research question addressed in this chapter is motivated by the economics literature related to pro-social behavior. A set of recent field experiments shows that financial and non-financial incentives have a positive effect on social behavior (Ashraf et al., 2014; Miller et al., 2012; Olken et al., 2012). These experiments are all conducted in settings in which social behavior (related to health and educational outcomes) is the main aim of the program and thus the main outcome variable of interest. However, social behavior has also become a topic of interest for many (large) for-profit firms, and social goals are sometimes added to traditional financial goals (Kitzmueller and Shimshack, 2012). Within these companies a method has to be found to effectively balance these two (possibly conflicting) goals. A potential solution to induce sustainable behavior could be through the use of incentives. Various studies have looked at the association between incentives and corporate social performance, and the empirical evidence is rather mixed (see Walls et al. (2012) for an overview). The drawback of many empirical studies is that they do not account for a non-random, i.e., endogenous, matching between a CEO and a firm, and between a firm and its remuneration policy. Hence, it is virtually impossible to establish a causal link when studying the association between CEO incentives and CSR using only observational data. The experimental design described in Chapter 5 solves these methodological difficulties while studying the effect of incentives on social behavior in a setting in which financial performance also matters.

To this end, the schools participating in an entrepreneurship education program are randomly assigned to one of three treatments: the first is purely financially oriented, the second promotes sustainable behavior and the third also induces sustainability by incentives. We compare the sustainable behavior of the teams in the three treatment groups and we find that (financial) incentives have a significant positive effect on sus-

tainable behavior. There is no difference in the financial performance between the teams in the different treatments. These results imply that, in this setting, the choice to behave more sustainable is not associated with financial performance. Moreover, the two sustainability treatments (with and without incentives) seem to have a positive effect on the individual environmental attitudes of the children in these treatments. These results suggest that even though discussing and encouraging CSR does not lead to a direct behavioral change, it does have a positive influence on the children's attitude towards the environment.

**Experimental setting**

The link between the three chapters is primarily given by the use of the same experimental setting and the same type of research method. The field experiments described in this dissertation were conducted within the setting of one of the leading, internationally renowned entrepreneurship education programs for primary schools (called BizWorld). The BizWorld program aims to teach children aged 11 or 12 the basics of business and entrepreneurship through an experiential learning program that takes five days (within a time span of 2 to 4 weeks). At the start of the program, the class is divided into teams of five or six children. Within each team, each team fulfills his/her specific role (e.g. CEO, CFO, Sales director, etc.) besides working together as a team. During the lessons, all five with a practical orientation, the children set up a toy business in friendship bracelets and go through a firm's entire business cycle (from start-up to liquidation). More specifically, the teams have to: write and present a business plan in order to raise start-up capital, design and manufacture products (friendship bracelets), calculate production costs and determine product prices, sell the products during a sales market to the pupils in the grade below, and finally they have to complete a profit and loss statement. Individual team members have strong incentives to care about the business performance of their team. In the school year 2009-2010, when we conducted the field experiments on team composition and sustainability, team members of the winning team were awarded a gift voucher of €7,50 each, and the team members of the runner up were each awarded a gift voucher of €5,00. This is an addition to the certificate for the winning team that is provided by the BizWorld foundation.

**Research method and contribution**

This dissertation reports the results from three field experiments. Field experiments are often used in medical trials, but up until recently their application in economics was scarce. In this dissertation this research method is applied to address the three research questions described above.

The aim of medical trials is typically to estimate a causal effect of a new drug on the treatment of a certain disease. Similarly, the field experiments described in this dissertation are also aimed at estimating the causal effect of a certain treatment variation on the outcome variables of interest. To be able to estimate a causal effect, a group of people (or subjects) is *randomly* divided into two or more subgroups, each with a certain treatment. As a result of the random treatment assignment, the subjects in the different groups are on average equal in terms of observed *and* unobserved characteristics. Hence, the only difference between these groups is in the treatment that they are exposed to. Thus, any differences that are observed in the average group outcomes can be attributed directly to the treatment.

In medical trials the treatment typically consists of the old medicine, the new medicine and a placebo, which serves as the control group. In this dissertation the treatments and the treatment groups vary in each chapter. Chapter 3 evaluates the effectiveness of the entrepreneurship education program. Thus, in this setting, the treatment is the entrepreneurship program itself and the control group (or the placebo) is the regular school curriculum. The outcome variables of interest in this chapter are the development of entrepreneurial skills and knowledge. The experiment described in Chapter 4 is about how team composition (in terms of skill balance) affects team performance. In this chapter the treatment thus consists of different team compositions, i.e., children are randomly assigned to a certain team conditional on their individual skill set. The outcome variable of interest in Chapter 4 is team performance. Chapter 5 tests the effect of incentives on sustainable behavior. The chapter reports the results of two treatments that are aimed at inducing this type of behavior. The outcome variable of interest is sustainable behavior and the teams in the regular course setting serve as the control group.

In addition to the estimation of causal effects, each chapter contributes to the economics literature in other ways. Chapter 3 is the first study to evaluate the effects of entrepreneurship education of children in primary school (ages 11 and 12). Previous studies of the impact of youth entrepreneurship education follow adolescents and mainly focused entrepreneurial intentions. Thus, another contribution of this chapter is the focus on the development of both knowledge and skills. The contribution of Chapter 4 is based on the fact that thus far there is little evidence on the effect of (balanced) skills at the team level on (team) performance. Another contribution of this chapter is that the experimental design creates the opportunity to explicitly study the intra-team substitutability of useful combinations of skills. Chapter 5 uses an experimental design to study sustainable behavior in a productive environment in which both sustainable behavior and financial performance matter. Furthermore, within this setting the aim is to study sustainable behavior at the team level, instead of at the individual level.

# Samenvatting (Summary in Dutch)

Het doel van dit proefschrift is om antwoord te geven op drie specifieke onderzoeksvragen: (1) Kun je ondernemerschap leren (op school, als je jong bent)?, (2) Hoe beïnvloeden verschillende combinaties aan kennis en vaardigheden binnen een team de prestaties van het team? en (3) Kunnen externe prikkels duurzaam gedrag bevorderen?

### Onderzoeksvraag 1

Hoofdstuk 3 is gebaseerd op het paper "The effect of early entrepreneurship education: Evidence from a field experiment", met als co-auteurs Randolph Sloof en Mirjam van Praag (European Economic Review 72, 2014, 76–97). Het doel van dit hoofdstuk is om de effectiviteit van ondernemerschapslesprogramma's op jonge leeftijd te onderzoeken. Het onderzochte lesprogramma BizWorld wordt wereldwijd gebruikt om kennis en vaardigheden op het gebied van ondernemerschap te ontwikkelen bij leerlingen in groep 8 van de basisschool.

Of je ondernemerschap kunt leren is een veelbesproken vraag, zowel in de academische wereld (bijv. Colombier and Masclet, 2008; Lindquist et al., 2013) als ook daarbuiten. De sterke toename in het aantal ondernemerschapslesprogramma's lijkt te suggereren dat men er in het algemeen van uitgaat dat je ondernemerschap kunt aanleren. Gezien het belang van ondernemerschap voor het stimuleren van economische groei en werkgelegenheid is het voor beleidsmakers een interessante gedachte dat ondernemerschap niet alleen is aangeboren, maar dat het ook kan worden aangeleerd. Dit zorgt er namelijk voor dat je ondernemerschap kunt stimuleren door middel van lesprogramma's. Tot nu toe is de effectiviteit van dit soort lesprogramma's echter nog maar mondjesmaat onderzocht. In hoofdstuk 3 van dit proefschrift bestudeer ik de effectiviteit van een ondernemerschapslesprogramma voor kinderen in groep 8 op de basisschool, dus voor kinderen op (relatief) jonge leeftijd. De motivatie om te kijken naar het investeren in vaardigheden op jonge leeftijd komt voort uit recent economisch onderzoek (Cunha and Heckman, 2007). Dit onderzoek toont aan dat een investering in kennis en vaardigheden op jonge leeftijd niet alleen een direct effect heeft op deze kennis en vaardigheden, maar ook zorgt voor positieve *spillovers* in de toekomst door als het ware een vruchtbare bodem te creëren waardoor toekomstige investeringen meer

opleveren. Vroeg investeren in vaardigheden kan op die manier dus een positief effect hebben op de lange termijn. Uiteraard leveren investeringen alleen deze (mogelijke) toekomstige voordelen op als een lesprogramma in de eerste plaats al een direct effect heeft op het ontwikkelen van de beoogde kennis en vaardigheden (op korte termijn). In hoofdstuk 3 van dit proefschrift onderzoek ik dan ook de directe (kortetermijn)effecten van dit ondernemerschapslesprogramma.

De resultaten in dit hoofdstuk laten zien dat het ondernemerschapsonderwijs een positief effect heeft op de ondernemerschapsvaardigheden van de aan het lesprogramma deelnemende leerlingen. Zelfvertrouwen, prestatiegerichtheid, risicobereidheid, analytisch vermogen, doorzettingsvermogen, pro-activiteit en creativiteit namen significant toe. Dit is opmerkelijk aangezien eerder onderzoek naar de effecten van ondernemerschapsonderwijs onder leerlingen en studenten in het voortgezet en hoger onderwijs, geen of soms zelfs negatieve effecten laten zien. Los van elkaar zijn deze vaardigheden belangrijk voor zowel ondernemers als werknemers. Onderzoek heeft echter aangetoond dat de combinatie van deze vaardigheden een positief effect heeft op het succes van ondernemers. De basisschool lijkt daarmee een vruchtbare bodem te bieden voor het ontwikkelen van deze vaardigheden. Vroeg investeren in deze vaardigheden is wellicht belangrijker dan tot nu toe werd gedacht.

**Onderzoeksvraag 2**

Hoofdstuk 4 is gebaseerd op "Jacks-of-all-trades? The effect of balanced skills on team performance", met als co-auteurs Randolph Sloof en Mirjam van Praag. Het doel van dit hoofdstuk is om te onderzoeken hoe de combinatie van kennis en vaardigheden binnen een team de prestaties van het team beïnvloeden.

De onderzoeksvraag die wordt onderzocht in hoofdstuk 4 is gebaseerd op de volgende twee observaties. Ten eerste: taken worden binnen organisaties steeds vaker uitgevoerd in teamverband (Hamilton et al., 2003). Dit geldt niet alleen voor grote multinationals, maar ook nieuwe ondernemingen worden steeds vaker opgericht door een team van ondernemers (Lazear and Shaw, 2007; Klotz et al., 2014). Ten tweede: we zien ook dat taken steeds complexer worden, onder andere door snelle technologische ontwikkelingen en globalisering (Lazear, 1999; Dahlin et al., 2005). De toename in complexe taken zorgt voor een toenemende vraag naar specifieke kennis en vaardigheden. De combinatie van complexe taken (en de bijbehorende vraag naar kennis en vaardigheden) en de toename aan taken die worden uitgevoerd in teamverband vraagt om onderzoek naar hoe de samenstelling van teams de teamprestatie beïnvloedt. In hoofdstuk 4 van dit proefschrift probeer ik deze vraag te beantwoorden door te kijken naar de samenstelling van bepaalde vaardigheden binnen een team.

Dit onderzoek bouwt onder andere voort op eerder onderzoek van Rulke and Galask-

iewicz (2000) dat laat zien dat teams bestaande uit generalisten beter presteren dan teams die bestaan uit specialisten. Daarbij maken Rulke and Galaskiewicz (2000) echter geen onderscheid tussen homogene en heterogene specialistenteams, dus tussen teams die bestaan uit één type specialisten en teams die bestaan uit een combinatie van verschillende typen specialisten. In hoofdstuk 4 wordt expliciet onderscheid gemaakt tussen verschillende typen specialisten. Op basis van de (CITO) scores van de individuele leerlingen voor taal en rekenen stellen we in totaal 179 teams samen, verdeeld over vier typen: generalistenteams, rekenspecialistenteams, taalspecialistenteams en gemengde specialistenteams. De resultaten uit dit onderzoek laten zien dat generalistenteams beter presteren dan taalspecialistenteams en gemengde specialistenteams. Dit wijst erop dat evenwichtige kennis en vaardigheden binnen een team, zoals in de generalistenteams, een positief effect heeft op de teamprestatie. Het verschil in prestatie tussen de generalistenteams en de gemengde specialistenteams laat zien dat het lastig is om verschillende typen specialisten met elkaar te laten samenwerken om op die manier evenwichtige kennis en vaardigheden binnen een team te creëren.

**Onderzoeksvraag 3**

Hoofdstuk 5 is gebaseerd op het paper "The effect of incentives on sustainable behavior: Evidence from a field experiment", eveneens met als co-auteurs Randolph Sloof en Mirjam van Praag. Dit hoofdstuk onderzoekt hoe je duurzaam gedrag kunt bevorderen in een productieomgeving. Bovendien wordt er gekeken naar de relatie tussen duurzaam gedrag en financiële prestaties. Om dit te onderzoeken worden de scholen die deelnemen aan het BizWorld lesprogramma onderverdeeld in drie groepen. In de eerste groep wordt de prestatie en de beloning van de teams alleen bepaald op basis van financiële prestaties. In de tweede groep wordt er expliciet aandacht besteed aan duurzaamheid en wordt duurzaam gedrag aangemoedigd (maar niet beloond). In de derde groep wordt duurzaam gedrag niet alleen besproken en aangemoedigd maar ook expliciet beloond.

De onderzoeksvraag die wordt beantwoord in hoofdstuk 5 is geïnspireerd door de economische literatuur ten aanzien van sociaal gedrag. De bevindingen van een aantal onlangs uitgevoerde veldexperimenten laten zien dat zowel financiële als niet-financiële prikkels sociaal gedrag kunnen bevorderen. Deze studies onderzoeken sociaal gedrag (zoals bijvoorbeeld het verbeteren van onderwijs of gezondheid) bij non-profit organisaties, dus in situaties waar het sociale gedrag het belangrijkste doel is van de programma's die worden bestudeerd. Maar in de laatste jaren wordt sociaal gedrag ook steeds belangrijker bij grote (for-profit) organisaties (in de vorm van Corporate Social Responsibility (CSR)). Deze organisaties nemen dan vaak naast financiële doelstellingen ook sociale doelstellingen op, zoals doelstellingen die bijvoorbeeld te maken hebben met het milieu of de diversiteit van hun werknemers (Kitzmueller and Shimshack, 2012).

Het probleem is echter dat financiële doelstellingen niet altijd overeenkomen of samengaan met de sociale doelstellingen. Hierdoor moeten deze bedrijven een manier zien te vinden om ervoor te zorgen dat beide doelen worden nagestreefd. Een mogelijke manier om dit te bereiken is door middel van het geven van externe (financiële) prikkels: mensen belonen voor bepaald gedrag. Er is wel onderzoek gedaan naar de relatie tussen beloning en CSR, maar het verband is niet eenduidig (Walls et al., 2012). Een mogelijke verklaring hiervoor is dat veel studies niet (kunnen) controleren voor zogenaamde "endogeniteit" in de keuze van een CEO om voor een bepaald bedrijf te werken en in de keuze van het bedrijf voor het aangeboden beloningspakket. Endogeniteit houdt in dat deze keuze gecorreleerd is met onobserveerbare karakteristieken van de CEO en/of het bedrijf, die tegelijkertijd gecorreleerd zijn met de uitkomstvariabele. Hierdoor is het bijna onmogelijk om een causaal verband aan te tonen tussen bestuurdersbeloningen en CSR op basis van het observeren van het gedrag van bestaande CEO's en bedrijven. Het experiment dat ik heb uitgevoerd en dat wordt beschreven in hoofdstuk 5 lost deze methodologische problemen op. Daarnaast is het mogelijk om in deze experimentele setting sociaal gedrag te bestuderen in een omgeving waarin ook financiële prestaties van belang zijn.

Als we het duurzame gedrag van de drie bovengenoemde groepen vergelijken zien we dat het belonen van duurzaam gedrag een significant positief effect heeft op dit gedrag. Er is geen verschil in de financiële prestaties tussen de teams in de verschillende groepen, dus in dit geval heeft duurzaam gedrag geen negatieve gevolgen. Tenslotte zien we dat de leerlingen in de twee duurzaamheidsgroepen (met en zonder beloning) zich meer bewust zijn van het milieu dan de leerlingen in de controlegroep. Dit laat zien dat, ook al heeft het bespreken en aanmoedigen van duurzaamheid geen direct effect op duurzaam gedrag, het wel bijdraagt aan een positieve houding ten aanzien van het milieu.

**Onderzoeksopzet**

Het verband tussen de drie hoofdstukken is gebaseerd op de onderzoeksmethode en op de setting waarin deze drie experimenten zijn uitgevoerd. Voor de experimenten in alle drie de hoofstukken heb ik gebruik gemaakt van data over leerlingen die deelnemen aan het BizWorld lesprogramma in groep 8 van de basisschool in Nederland. Het doel van het lesprogramma is om kinderen op jonge leeftijd basisvaardigheden en kennis over ondernemerschap aan te leren. Het lesprogramma duurt vijf dagen die worden gegeven verspreid over een periode van twee tot vier weken. Aan het begin van het programma worden teams samengesteld met vijf tot zes leerlingen per team. Binnen het team heeft iedere leerling zijn eigen functie (zoals Algemeen directeur, Financieel directeur, directeur Marketing, etc.), maar uiteindelijk gaat het om de teamprestatie.

Gedurende het lesprogramma vormt ieder team een bedrijfje in vriendschapbandjes en doorlopen ze de verschillende fasen van een start-up. Zo moeten de teams onder andere een bedrijfsplan schrijven en dit plan presenteren aan een investeerder om aandelen te verkopen en startkapitaal te verkrijgen, producten (vriendschapsbandjes) ontwerpen en maken, productiekosten berekenen en de verkoopprijs bepalen, producten verkopen tijdens de 'Big Sale' aan de kinderen uit groep 7 en tenslotte moeten ze de financiële afrekening doen (winst-en-verliesrekening opmaken). Alle teams willen erg graag winnen en vinden het belangrijk dat het goed gaat met hun bedrijf. Om te zorgen voor extra motivatie konden de teams die meededen aan het veldexperiment in het voorjaar van 2010 (voor het onderzoek naar teamsamenstelling en duurzaamheid) ook nog cadeau- (of boeken)bonnen winnen. De kinderen uit het winnende team kregen ieder een bon ter waarde van €7,50, de kinderen uit het team dat tweede werd kregen ieder een bon ter waarde van €5,00.

**Methodiek**

In dit proefschrift beschrijf ik de resultaten van drie veldexperimenten. Veldexperimenten worden vaak gebruikt bij medisch onderzoek maar werden tot tien jaar geleden in de economische wetenschappen nog maar weinig toegepast. In dit proefschrift maak ik gebruik van deze onderzoeksmethode om antwoord te geven op de drie hiervoor genoemde economische onderzoeksvragen.

Het doel van medische experimenten is meestal het onderzoeken van het effect van een bepaald medicijn op de behandeling van een bepaalde ziekte. Op dezelfde manier worden de veldexperimenten in dit proefschrift gebruikt om een causaal (d.w.z. oorzakelijk) verband aan te tonen tussen een bepaalde "behandeling" en een uitkomstvariabele. Om een causaal verband te kunnen aantonen is het van belang dat een groep mensen willekeurig in twee of meer groepen wordt opgedeeld. Door deze willekeurige toewijzing zijn de mensen in iedere groep gemiddeld genomen vergelijkbaar, zowel in zichtbare als onzichtbare persoonskenmerken. De willekeurige indeling zorgt er dus voor dat de toegewezen "behandeling" het enige verschil is tussen de groepen. Alle verschillen in de gemiddelde groepsuitkomsten kunnen dus rechtstreeks aan de behandeling worden toegeschreven en zijn dus het directe effect van de behandeling.

Bij medisch onderzoek bestaat de behandeling vaak uit het oude medicijn, een nieuw medicijn en een placebo (de controlegroep). In dit proefschrift onderzoek ik in ieder hoofdstuk het effect van een andere "behandeling". Voor de evaluatie van het ondernemerschapslesprogramma in hoofdstuk 3 bestaat de behandeling uit het BizWorld lesprogramma zelf en de controlegroep (of de placebo) zijn de reguliere lessen gedurende die periode in groep 8. De uitkomstvariabelen waar ik geïnteresseerd in ben zijn 'kennis over ondernemerschap' en bepaalde ondernemerschapsvaardigheden. De

"behandeling" die ik onderzoek in hoofdstuk 4 is teamsamenstelling. Voor dit experiment worden kinderen willekeurig aan een team toegewezen op basis van hun eigen kennis en vaardigheden. In dit hoofstuk zijn de uitkomstvariabelen de teamprestaties tijdens het lesprogramma. Voor het onderzoek in hoofdstuk 5 bestaat de "behandeling" uit verschillende beloningsstructuren die ik willekeurig aan de deelnemende scholen heb toegewezen. De verschillende manieren van belonen zorgen voor verschillende prikkels om duurzaam gedrag te stimuleren. Om het effect van de verschillende prikkels te meten vergelijk ik het duurzame gedrag van de teams in de verschillende "behandelingen" (beloningsstructuren).

**Wetenschappelijke bijdrage**

De bijdrage van dit proefschrift bestaat niet alleen uit het schatten van causale effecten. Ieder hoofdstuk levert nog andere specifieke bijdragen aan de economische literatuur. Hoofdstuk 3 is het eerste onderzoek naar de effecten van ondernemerschapsonderwijs op jonge leeftijd. Eerder onderzoek was tot nu toe vooral gericht op studenten in het hoger onderwijs of een enkele keer op leerlingen in het voortgezet onderwijs. Daarnaast ligt de nadruk in dit hoofdstuk op de ontwikkeling van kennis en vaardigheden op het gebied van ondernemerschap (in tegenstelling tot eerdere studies die met name keken naar het voornemen om ondernemer te worden). Het onderzoek in hoofdstuk 4 draagt bij aan de kennis over de optimale samenstelling van teams wat betreft de combinatie van kennis en vaardigheden. Bovendien is het met dit experiment mogelijk om de substitueerbaarheid van bepaalde combinaties van vaardigheden binnen het team te onderzoeken. In hoofdstuk 5 bestudeer ik duurzaam gedrag in een productieomgeving waarin zowel duurzaam gedrag als de financiële prestaties van belang zijn. Daarnaast ligt de nadruk in dit hoofdstuk op duurzaam gedrag op teamniveau in plaats van op individueel niveau.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

**560** S.P. KASTORYANO, *Essays in Applied Dynamic Microeconometrics*

**561** M. VAN DUIJN, *Location, choice, cultural heritage and house prices*

**562** T. SALIMANS, *Essays in Likelihood-Based Computational Econometrics*

**563** P. SUN, *Tail Risk of Equidity Returns*

**564** C.G.J. KARSTEN, *The Law and Finance of M&A Contracts*

**565** C. OZGEN, *Impacts of Immigration and Cultural Diversity on Innovation and Economic Growth*

**566** R.S. SCHOLTE, *The interplay between early-life conditions, major events and health later in life*

**567** B.N.KRAMER, *Why don't they take a card? Essays on the demand for micro health insurance*

**568** M. KILIÇ, *Fundamental Insights in Power Futures Prices*

**569** A.G.B. DE VRIES, *Venture Capital: Relations with the Economy and Intellectual Property*

**570** E.M.F. VAN DEN BROEK, *Keeping up Appearances*

**571** K.T. MOORE, *A Tale of Risk: Essays on Financial Extremes*

**572** F.T. ZOUTMAN, *A Symphony of Redistributive Instruments*

**573** M.J. GERRITSE, *Policy Competition and the Spatial Economy*

**574** A. OPSCHOOR, *Understanding Financial Market Volatility*

**575** R.R. VAN LOON, *Tourism and the Economic Valuation of Cultural Heritage*

**576** I.L. LYUBIMOV, *Essays on Political Economy and Economic Development*

**577** A.A.F. GERRITSEN, *Essays in Optimal Government Policy*

**578** M.L. SCHOLTUS, *The Impact of High-Frequency Trading on Financial Markets*

**579** E. RAVIV, *Forecasting Financial and Macroeconomic Variables: Shrinkage, Dimension reduction, and Aggregation*

**580** J. TICHEM, *Altruism, Conformism, and Incentives in the Workplace*

**581** E.S. HENDRIKS, *Essays in Law and Economics*

**582** X. SHEN, *Essays on Empirical Asset Pricing*

**583** L.T. GATAREK, *Econometric Contributions to Financial Trading, Hedging and Risk Measurement*

**584** X. LI, *Temporary Price Deviation, Limited Attention and Information Acquisition in the Stock Market*

**585** Y. DAI, *Efficiency in Corporate Takeovers*

**586** S.L. VAN DER STER, *Approximate feasibility in real-time scheduling: Speeding up in order to meet deadlines*

**587** A. SELIM, *An Examination of Uncertainty from a Psychological and Economic Viewpoint*

**588** B.Z. YUESHEN, *Frictions in Modern Financial Markets and the Implications for Market Quality*

**589** D. VAN DOLDER, *Game Shows, Gambles, and Economic Behavior*

**590** S.P. CEYHAN, *Essays on Bayesian Analysis of Time Varying Economic Patterns*

**591** S. RENES, *Never the Single Measure*

**592** D.L. IN 'T VELD, *Complex Systems in Financial Economics: Applications to Interbank and Stock Markets*

**593** Y.YANG, *Laboratory Tests of Theories of Strategic Interaction*

**594** M.P. WOJTOWICZ, *Pricing Credits Derivatives and Credit Securitization*

**595** R.S. SAYAG, *Communication and Learning in Decision Making*

**596** S.L. BLAUW, *Well-to-do or doing well? Empirical studies of wellbeing and development*

**597** T.A. MAKAREWICZ, *Learning to Forecast: Genetic Algorithms and Experiments*

**598** P. ROBALO, *Understanding Political Behavior: Essays in Experimental Political Economy*

**599** R. ZOUTENBIER, *Work Motivation and Incentives in the Public Sector*

**600** M.B.W. KOBUS, *Economic Studies on Public Facility use*

**601** R.J.D. POTTER VAN LOON, *Modeling non-standard financial decision making*

**602** G. MESTERS, *Essays on Nonlinear Panel Time Series Models*

**603** S. GUBINS, *Information Technologies and Travel*

**604** D. KOPÁNYI, *Bounded Rationality and Learning in Market Competition*

**605** N. MARTYNOVA, *Incentives and Regulation in Banking*

**606** D.KARSTANJE, *Unraveling Dimensions: Commodity Futures Curves and Equity Liquidity*

**607** T.C.A.P. GOSENS, *The Value of Recreational Areas in Urban Regions*

**608** Ł.M. MARĆ, *The Impact of Aid on Total Government Expenditures*

**609** C.LI, *Hitchhiking on the Road of Decision Making under Uncertainty*