



Paper to be presented at the  
DRUID Society Conference 2014, CBS, Copenhagen, June 16-18

## **Ability Dispersion and Team Performance**

**Sander Hoogendoorn**  
University of Amsterdam  
Amsterdam School of Economics  
sander\_hoogendoorn@yahoo.com

**Simon Parker**  
University of Western Ontario  
Ivey Business School  
sparker@ivey.uwo.ca

**Mirjam Van Praag**  
Copenhagen Business School  
Dept of Innovation and Organizational Economics  
mvp.ino@cbs.dk

### **Abstract**

What is the effect of dispersed levels of cognitive ability of members of a (business) team on their team's performance? This paper reports the results of a field experiment in which 573 students in 49 teams start up and manage real companies under identical circumstances for one year. We ensured exogenous variation in -otherwise random- team composition by assigning students to teams based on their measured cognitive abilities (Raven test). Each team performs a variety of tasks, often involving complex decision making. The key result of the experiment is that the performance of business teams first increases and then decreases with ability dispersion. We seek to understand this finding by developing a model in which team members of different ability levels form sub-teams with other team members with similar ability levels to specialize in different productive tasks. Diversity spreads production over different tasks in order to escape diminishing marginal returns under specialization. The model comes with a boundary condition: our experimental finding is most likely to emerge in settings where different tasks exhibit moderate differences in their productive contributions to total output.

# Ability Dispersion and Team Performance

## **Abstract**

What is the effect of dispersed levels of cognitive ability of members of a (business) team on their team's performance? This paper reports the results of a field experiment in which 573 students in 49 teams start up and manage real companies under identical circumstances. We ensured exogenous variation in — otherwise random — team composition by assigning students to teams based on their measured cognitive abilities (Raven test). Each team performs a variety of tasks, often involving complex decision making. The key result of the experiment is that the performance of business teams first increases and then decreases with ability dispersion. We seek to understand this finding by developing a model in which team members of different ability levels form sub-teams with other team members with similar ability levels to specialize in different productive tasks. Diversity spreads production over different tasks in order to escape diminishing marginal returns under specialization. The model comes with a boundary condition: our experimental finding is most likely to emerge in settings where different tasks exhibit moderate differences in their productive contributions to total output.

JEL-codes: C93, D83, J24, L25, L26, M13, M54

Keywords: Ability dispersion, team performance, field experiment, entrepreneurship

## 1 Introduction

Decision-making in organizations is increasingly performed by teams rather than by individuals (Hamilton et al., 2003; Woolley et al., 2010). Consequently, the optimal composition of teams in terms of member abilities can be regarded as a valuable organizational asset. Yet, despite it being widely believed that (cognitive) abilities of members affect overall team performance, the precise impact of ability dispersion on team performance remains poorly understood (Hamilton et al., 2012). Specifically, we lack evidence about whether cognitive ability diversity of team members is or is not conducive to the performance of teams. Such evidence is potentially useful for managers seeking to select members of internal work teams, as well as being of interest in its own right. Indeed, teams often exert considerable influence on the performance of public and private organizations since they take decisions of strategic and operational importance.

This paper examines the effect of dispersion in cognitive ability (hereafter, just 'ability') on the performance of business teams. We do so by creating a field experiment in which teams of undergraduate students start up and manage a real company as a compulsory part of the curriculum in an international business program in the Netherlands. Companies are simultaneously founded on a level-playing field; and students face strong incentives that align their interests with the business performance of their company. Our experiment randomizes 573 students into 49 teams conditional on their measured cognitive abilities. We ensure a relatively large exogenous variation in ability dispersion between teams to help probe non-linearities in the relationship between ability dispersion and team performance.

We believe there are two principal advantages of our empirical design. First, a field experiment can establish a causal relationship between team composition and performance, in contrast to observational studies in which members are free to self-select into and out of teams (Hansen et al., 2006). Self-selection confounds the identification of 'treatment' effects of team composition, a problem that our field experiment is designed to overcome. Second, relative to prior empirical studies which have analyzed settings involving laboratory experiments (Woolley et al., 2010), unskilled work tasks (e.g., Hamilton et al., 2003), or competitive sports (Kahn, 2000), our field experiment closely resembles the functioning of teams co-operating on a complicated real business project. The tasks of these teams are often complex and broad in scope, entailing the sustained application of

members' cognitive abilities.

The findings of this experiment can be summarized as follows. Team performance as measured in terms of sales, profits and profits per share first increases at low levels of ability dispersion up to a maximum before decreasing at higher levels of dispersion. Controlling for the average cognitive ability of teams, performance is maximized at a coefficient of variation in cognitive ability of approximately 0.25 (the sample average is 0.22). That is, firm performance displays an inverse-U shaped relationship with the diversity in a team in terms of its members' cognitive ability levels.

Prior theorizing turns out to be unable to explain these results. Previous theories about ability diversity in teams have highlighted benefits to, or costs from, diversity, but not both at the same time — predicting monotonic relationships between member ability diversity and team performance. So, for example, if the inputs of members are complementary in team production, performance is maximized by matching individuals of similar abilities (Prat, 2002). Conversely, if a few able team members can impose high team production norms, and/or enable learning by less able members, diverse teams can be best (Hamilton et al., 2003). These arguments have been applied to explain performance in team sports — see, e.g., Gould and Winter (2009) in the case of Major League Baseball, and Franck and Nüesch (2010) in the case of German professional soccer — and routinized factory production (Hamilton et al., 2003). However, these models may be less applicable in business environments where performance depends on the outputs of multiple tasks.

In the light of these limitations, we make an additional contribution by proposing a model that reflects salient features of the experimental design while also generating predictions consistent with the empirical findings. Specifically, the model posits a multi-task environment where teams choose how many tasks they wish to perform, and seek the most effective ways of performing each task (in a set of 'sub-teams') to create value. We show that diversity may be advantageous in this multi-task environment because it spreads production over different tasks in order to escape diminishing marginal returns under specialization. However, excessive diversity may concentrate too much production in less productive tasks which harms performance. The main result from the model is the following proposition: expected performance is strictly increasing (decreasing) in team diversity when productivity differences between different tasks are high (low). With intermediate productivity differences between tasks, firm performance displays an inverse-U shaped relationship with team diversity — consistent with the findings of our experiment.

The theoretical study that is most related to ours is Hong and Page (2004). In their model, teams have to solve a single task; and people of different abilities choose different solution algorithms. Hong and Page show that as the number of team members increases, the ablest members necessarily become similar in the space of problem solvers. That limits the performance of homogeneous teams even if they are comprised of the ablest people. Diverse teams, in contrast, enjoy the benefit of searching more of the solution space, and so can outperform homogeneous teams with higher average ability. While our model also incorporates the idea that team members search for the best ways of executing tasks, we generalize the setting to multiple tasks with different productivity levels. We do not predetermine our results by assuming a particular team production function, using instead a neutral additive structure to make our point most clearly.

The remainder of this paper is organized as follows. Section 2 outlines the context, design and data of our field experiment. Section 3 presents the empirical findings. Section 4 proposes a simple theoretical model which can potentially explain the results. Section 5 provides a brief discussion and conclusion.

## 2 Field experiment

### 2.1 Context

The teams in our field experiment are teams of undergraduate students that have to start up and manage a real company as a compulsory part of the curriculum at the Department of International Business Studies of the Amsterdam College of Applied Sciences.<sup>1</sup> The entrepreneurship program covers about one-fifth of students' first-year undergraduate curriculum. This program is organized in collaboration with Junior Achievement (JA), which is the worldwide leading provider of educational programs in entrepreneurship (Oosterbeek et al., 2010). Companies in the entrepreneurship program are simultaneously founded on a level playing field and dissolved after one academic year. The experiment was performed in 2009-2010 and its context and incentives are similar to those described in Hoogendoorn and Van Praag (2012) and Hoogendoorn et al. (2013), that were performed at the same college one year before.

---

<sup>1</sup>This department is subdivided into five sub-departments/fields of study: business management, management, trade management Asia, business languages and financial management. Students are assigned to teams within these sub-departments/fields of study. This does not affect the randomization process, as explained below.

During the program students: raise capital by issuing shares; appoint officers and delegate tasks; produce and market products or services; keep the accounts; and conduct shareholders' meetings. Hence, students execute substantial and genuinely collaborative tasks that require them to establish roles, build up relationships, and create routines and processes in order to maximize shareholder value. Moreover, students face strong incentives that align their interests with the business performance of the company (see subsection 2.2 below). Each company reports to their randomly assigned professor and business coach on a regular basis. Everything about the company is real, including tax and social security payments. The program is not a business simulation. In sum, students in our experiment have to coordinate on a broad array of complex decision-making tasks that entail the sustained application of their cognitive abilities.

Companies typically proceed as follows. They start with brainstorming about potential business activities and conducting market research to select the most viable idea. There are no restrictions on the type of business activity that can be chosen. Simultaneously, teams appoint about half of their members to management positions (such as the CEO and CFO) and the other half to non-management positions, where management positions are redistributed among the non-managing part of the team halfway the program.<sup>2</sup> Companies further develop their chosen idea by writing a business plan, and they start raising capital by issuing shares. Other sources of financing such as personal or outside loans are not allowed. Once the business plan is authorized by the majority of shareholders at the first shareholders' meeting, business operations of teams boil down to production and marketing of the chosen products or services. All companies are dissolved at the end of the program and each team has to write an annual report that needs approval at the final shareholders' meeting. Any profits are divided among the shareholders.

The college prescribes that the 573 students set up 49 team-based companies. They allowed us to measure students' scores in the 20-minute timed version of Raven's advanced progressive matrices test as a proxy for cognitive ability (Hamel and Schmittmann, 2006; Raven et al., 1998) and assign students randomly to teams conditional on these scores. The average team size is equal to approximately 12 students. Table A1 in the appendix lists the key characteristics of all 49 teams,

---

<sup>2</sup>The ability composition of the entire team closely resembles that of the management team, which is possibly the more influential part of a team, although students rather tend to view their teams a unit and do not strongly distinguish between the managing and non-managing part of the team. Point estimates from a regression of the ability composition of the management team on the ability composition of the entire team are not significantly different from 1 before and after roles are switched.

including the product or service they market.

### *Cognitive ability*

Raven’s advanced progressive matrices are extensively used to differentiate between people of higher cognitive ability (Bors and Stokes, 1998; Mills et al., 1993; Raven et al., 1993). The test requires subjects to select the missing figure out of eight possibilities that completes a logical pattern (see Figure A1 in the appendix for an example). Patterns become increasingly difficult as subjects progress. Over the past decades, Raven’s advanced progressive matrices have been shown to associate with cognitive ability or intelligence in various ways. Elaborating on Spearman’s notion of general cognitive ability, Raven’s advanced progressive matrices are found to measure fluid intelligence (Cattell, 1963), analytic intelligence (Carpenter et al., 1990), and intellectual efficiency if administered with a time limit (Hamel and Schmittmann, 2006).<sup>3</sup> Test scores on Raven’s advanced progressive matrices can be interpreted as a proxy for cognitive ability. Indeed, the correlation between these test scores and students’ grade point average (GPA) shows a significant and positive relationship in our sample.

Table 1 provides descriptive statistics of cognitive ability at the team level (panel A) and by field of study (panel B). Panel A indicates that the average number of figures correctly solved in the 20-minute timed version of Raven’s advanced progressive matrices test is 18.60 out of 36 figures at maximum. We do not transform these test scores into an intelligence quotient, because that would require an additional assumption about the proper norm for first-year college students (Hamel and Schmittmann, 2006).<sup>4</sup> Moreover, our main interest is in the exogenous variation in cognitive ability rather than the exact level of test scores. Consistent with recent empirical studies involving professional sports (Franck and Nüesch, 2010; Papps et al., 2011), we use the coefficient of

---

<sup>3</sup>Spearman (1927) decomposed general cognitive ability ( $g$ ) into an eductive and a reproductive component, where (i) eductive ability reflects “the ability to make meaning out of confusion, the ability to generate high-level, usually nonverbal, schemata which make it easy to handle complexity”, and (ii) reproductive ability reflects “the ability to absorb, recall, and reproduce information that has been made explicit and communicated from one person to another” (Raven, 2000, p. 2). Fluid intelligence, analytic intelligence, intellectual efficiency and, hence, the 20-minute timed version of Raven’s advanced progressive matrices test, mainly relate to eductive ability. Nevertheless, scores on cognitive tests such as Raven’s advanced progressive matrices may differ across time, gender and culture (see Calvin et al., 2011; Irwing and Lynn, 2005; Nisbett et al., 2012; Raven, 2000; Rushton and Jensen, 2005).

<sup>4</sup>In line with Bors and Stokes (1998) and Raven et al. (1998) we exclude students with a test score  $\leq 6$  from the sample. Including this group of 10 students in total may incorrectly inflate teams’ ability dispersion since a test score  $\leq 6$ , in more conventional intelligence terms, roughly corresponds with the cognitive ability level of elementary school dropouts (which is highly unlikely for first-year college students). Students with a test score  $\leq 6$  most likely just did not put in effort or choked while taking the test. T-tests acknowledge that those students are not significantly different from students with a test score  $> 6$  in terms of age, gender, risk aversion and GPA.

**Table 1.** Descriptive statistics of cognitive ability

<b>A: Cognitive ability</b>								
	<u>Mean</u>	<u>SD</u>	<u>Min</u>	<u>Max</u>				
Average ability	18.60	2.53	14.00	23.22				
Ability dispersion (CV)	0.22	0.09	0.07	0.47				
<b>B: Field of study</b>								
	<u>Stud.</u>	<u>Teams</u>	<u>Average ability</u>			<u>Ability dispersion</u>		
			<u>Mean</u>	<u>Min</u>	<u>Max</u>	<u>Mean</u>	<u>Min</u>	<u>Max</u>
Business management	265	21	18.74	14.07	22.40	0.22	0.07	0.47
Management	45	4	18.67	16.27	22.22	0.23	0.18	0.33
Trade management Asia	108	10	18.92	14.78	23.22	0.20	0.10	0.35
Business languages	123	12	17.94	14.00	21.86	0.21	0.08	0.36
Financial management	32	2	19.33	18.83	19.83	0.27	0.21	0.32
Total	573	49	18.60	14.00	23.22	0.22	0.07	0.47

*Note:* Average and CV of ability reflect at the team level respectively average score and coefficient of variation in scores on Raven's advanced progressive matrices test.

variation in test scores as a scale-invariant measure for ability dispersion in teams. Teams' coefficient of variation in test scores varies between 0.07 and 0.47 with a sample average of approximately 0.22. Panel B shows the numbers of students and teams by field of study. It also indicates that (the range of) average ability and ability dispersion of teams are similar across fields of study; possibly except for the field of financial management which accommodates only two teams.

## 2.2 Design

The cognitive ability of students and their background characteristics were administered one week before the start of the entrepreneurship program.<sup>5</sup> As outside researchers we then manipulated the ability composition of teams and randomly assigned students to teams in accordance with our imposed variation in cognitive ability. In practice, we proceeded as follows.

Within fields of study, students were divided into four quartiles per class on the basis of their test score, where 1 reflects the best quartile and 4 the worst quartile. Each class was then split up in two teams, either combining cognitive ability quartiles 1+2 and 3+4 or, alternatively combining cognitive ability quartiles 1+4 and 2+3 in a class. Hence, '1+2 teams' have a high average ability and a low ability dispersion, '3+4 teams' have a low average ability and a low ability dispersion, '1+4 teams' have a medium average ability and a high ability dispersion, and '2+3 teams' have a

<sup>5</sup>Students were kept uninformed about their score in the 20-minute timed version of Raven's advanced progressive matrices test. We presented the fact that they were tested as a standard procedure of the introductory week at their new college.



medium average ability and a low ability dispersion. The assignment of students was implemented one week later by the program coordinators, who were informed about the character of our field experiment.<sup>6</sup> Students and business coaches were uninformed, whereas professors only knew that a research project was performed that prohibited students to switch teams. Only 6 out of 573 students managed to switch teams during the program.

Figure 1 shows frequency distributions of scores on Raven's advanced progressive matrices test at the individual and team level (average ability and ability dispersion). At the individual level test scores range from 7 to 32 figures correctly solved. The average ability of teams varies between test scores of 14 and 23, while as mentioned before teams' ability dispersion ranges from 0.07 to a coefficient of variation in test scores of 0.47. We exploit this substantial and exogenous variation in cognitive ability to study the impact of teams' heterogeneity in cognitive ability.

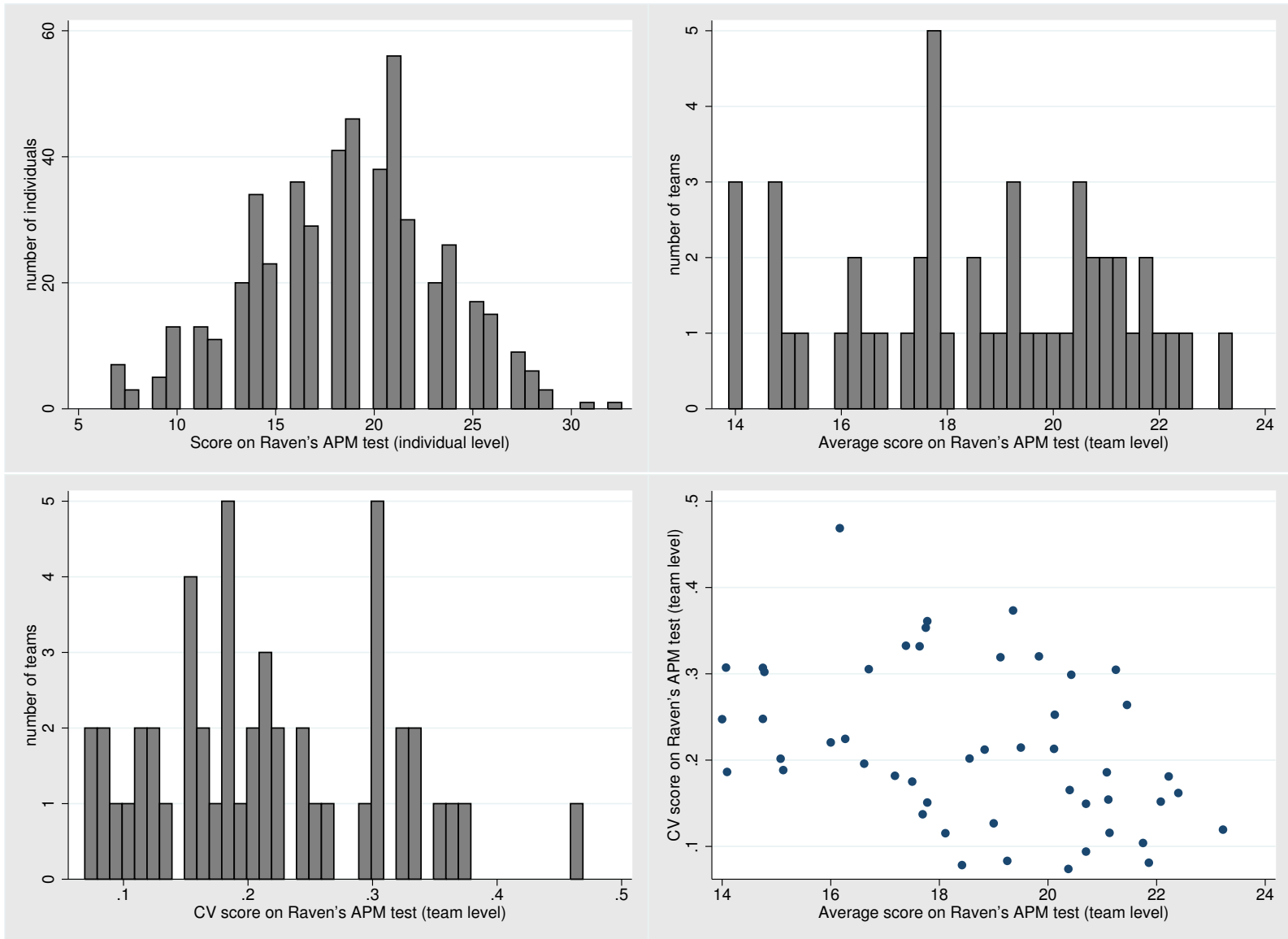
One might worry that the effect of ability dispersion on team performance is biased since teams of low or high average ability, by construction, tend to have a lower ability dispersion (relative to teams of medium average ability) because they have been drawn from a truncated distribution of ability. The scatter plot of teams' average ability and ability dispersion, however, does not reveal a systematic pattern that may confound a causal interpretation of the effect of ability dispersion on the business performance of teams (see Figure 1). Moreover, the results in section 3 are similar if we include only 31 medium-ability teams with average test scores not more than one standard deviation away from the average test score in the sample, i.e., with average test scores in the range of  $18.60 \pm 2.53$  (see Table 1).

Dropout rates for first-year students in Dutch higher vocational schools, where the admission of students based on grades or previous achievements is prohibited, are on average 30% including students that switch study and/or school (HBO-raad, 2010). The design of our experiment could be compromised if dropouts change the ability composition of teams. During the entrepreneurship program approximately 14% of the students dropped out (or were dismissed), which reduced the average team size from 12 to 10 students.<sup>7</sup> Nevertheless, this barely changed teams' overall average ability and ability dispersion. The average ability of teams increased from a test score of 18.60 to

---

<sup>6</sup>A few late applicants were randomly distributed among the existing teams whereas a few 'no shows' were also randomly distributed across teams (as they did not know to which team they were assigned to at that stage).

<sup>7</sup>Lower dropout rates than the national average at the Department of International Business Studies of the Amsterdam College of Applied Sciences can be explained by the fact that international programs generally attract more motivated students.



**Figure 1.** Frequencies and scatter of scores on Raven's advanced progressive matrices test (individual and team level; average and CV score)

18.65, while teams' ability dispersion remained the same at a coefficient of variation in test scores of 0.22. The correlation between teams' average ability at the start and end of the program is 0.96; the same correlation was observed for ability dispersion at the start and end of the program. Students in teams of different ability composition also did not drop out more or less often (dismissals will be discussed below). In sum, we are confident that these composition dynamics did not compromise the design of our experiment.

### *Incentives*

There are various strong incentives in place that align the interests of students with the business performance of their company. First and foremost, students can be dismissed in case of repetitive free-riding.<sup>8</sup> Dismissal of team members requires a two-third voting majority in the team together with the consent of the professor. It is a credible threat since the average number of dismissals is 0.35 per team and nearly 30% of the teams experiences at least one dismissal during the entrepreneurship program. Dismissal has severe consequences. Students are excluded from the program, lose its corresponding 12 credit points (out of 60 credit points in the first year) and endanger their prospect of obtaining an undergraduate degree (for which a minimum number of 45 credit points in the first year is mandatory). In section 3 we will examine to what extent dismissals vary across teams of different ability composition and show that dismissals do not endanger the design of the field experiment by affecting the ability composition of teams.

Another incentive is provided by the grade students obtain for the program from their professor, which has a substantial weight of 20% in their (first-year) grade point average. Both individual and team performance determine the program grade and their weight in the total program grade is about 50/50. Assessment of both components is based on the professor's subjective evaluation. Individual performance of students mainly entails active participation and professors' perceptions of the development of competencies such as cooperation, entrepreneurial behavior and professionalism. An indicator of the within team variation in individual performance is the average difference between the highest and lowest program grade within a team. This was 1.5 on a 10-point scale (s.d. 0.8). The relevance of team performance for the program grade is indicated by a significant and positive correlation between teams' average program grade and business outcomes.

---

<sup>8</sup>Interviews with program coordinators acknowledge that the main cause of dismissals is shirking of team members.

Virtually all students own one or more shares in their companies (with a nominal value of 20 euros per share). Roughly half of the shares are owned by team members themselves; the remaining shareholders are usually family members, friends and/or acquaintances. The mean number of shares issued is 57 (s.d. 26.1), while the minimum and maximum numbers of shares sold are 21 and 135. Finally, teams participate in a formal business competition. Six selected teams present their results in a ‘business pitch’ at the end of the program to a jury of entrepreneurs who choose a winner based on business outcomes and presentations. The winning team obtains a cup, often gets some (local) press attention and represents the college in a national competition. All in all, the incentives discussed above help ensure that students care about the business performance of their company.

### *2.3 Data*

We accessed various data sources to collect information about individuals and teams. One week before the start of the entrepreneurship program students took the 20-minute timed version of Raven’s advanced progressive matrices test and filled out a pretreatment questionnaire that mainly covered their background characteristics (response rate: 89%). Simultaneously, we received administrative data to assist us in assigning students to teams. At the end of the program, students filled out a posttreatment questionnaire that queried team characteristics and processes (response rate: 68%). We then also obtained the approved annual reports, which contain information about the business performance of teams (response rate: 100%). The data that we collected were used to: construct exogenous variation in cognitive ability across teams (see subsection 2.2); test the predictions of our model (see section 3); and assess whether the assignment of students to teams was random conditional on their cognitive ability (see below).

Business performance is operationalized by four measures: sales, profits, a binary indicator for positive profits and profits per share. We include a binary indicator for positive profits to account for the fact that many students view as the bottom line result whether or not they are able to satisfy shareholders. Table 2 shows that average sales for all 49 teams are equal to 902 euros and that profits are 24 euros on average. More than half of the teams makes a profit (57%) and average profits per share amount to 0.62 euros. All three profit measures are significantly and positively correlated with sales.

If we split the sample into teams of low ( $\text{mean} < 17$ ), moderate ( $17 \geq \text{mean} \leq 21$ ) and high ( $\text{mean} > 21$ )

**Table 2.** Descriptive statistics of business outcomes

		Average ability				Ability dispersion		
		All	Low mean<17	Moderate 17≥mean≤21	High mean>21	Low CV<0.15	Moderate 0.15≥CV≤0.30	High CV>0.30
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sales (euros)	Mean	902	939	944	762	714	1140	636
	SD	974	909	1111	756	537	1267	478
	Min	0	43	0	42	210	42	0
	Max	5049	3505	5049	2553	2019	5049	1576
Profits (euros)	Mean	24	-18	91	-80	-94	115	-37
	SD	422	347	492	323	391	491	273
	Min	-993	-637	-993	-592	-993	-637	-596
	Max	1619	413	1619	353	508	1619	376
Positive profits (0/1)	Mean	0.57	0.62	0.60	0.45	0.42	0.63	0.62
	SD	0.50	0.51	0.50	0.52	0.51	0.49	0.51
	Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Profits per share (euros)	Mean	0.62	0.37	1.43	-0.91	-0.85	1.94	-0.46
	SD	7.01	5.71	7.88	6.58	5.31	8.42	5.27
	Min	-13.76	-9.65	-10.13	-13.76	-10.13	-13.76	-8.25
	Max	25.70	9.89	25.70	8.02	8.02	25.70	9.89
N		49	13	25	11	12	24	13

*Note:* Average and CV of ability reflect at the team level respectively average score and coefficient of variation in scores on Raven's advanced progressive matrices test.

average ability, descriptive statistics suggest that teams of moderate average ability perform slightly better on the different business outcomes. However, note that these descriptives are very sensitive to the exact location of particularly the second cutoff point. A cutoff at a mean test score of 20 (instead of 21), for example, would already imply that teams of high average ability achieve better results. If we split up the sample in teams of low ( $CV < 0.15$ ), moderate ( $0.15 \geq CV \leq 0.30$ ) and high ( $CV > 0.30$ ) ability dispersion, teams of moderate ability dispersion tend to have higher sales, profits and profits per share than teams in the other two categories, on average. This ranking is rather insensitive to the precise location of the cutoff points. The effect of the average level and dispersion of ability in teams will be examined more formally in section 3.

### *Randomization*

To assess whether the assignment of students to teams was truly random (conditional on their cognitive ability), we regress background characteristics of students on the average test score in their team, the team's coefficient of variation in test scores, and its square — separately for low ability and high ability students. The selection of these three independent variables is consistent with the team level specifications of the main results in section 3, i.e., in combination they represent the main measures of a team's ability composition.

Panel A1 of Table 3 shows that background characteristics of low ability students do not systematically vary across teams of different ability composition. Hence, low ability students in teams of low ability dispersion are not significantly different from low ability students in teams of high ability dispersion. The same holds for background characteristics of high ability students (see panel A2). Low ability and high ability students assigned to teams of distinct ability composition are also not more or less likely to follow a specific field of study (not tabulated).

In a similar fashion, panel B of Table 3 examines at the team level whether (average) background characteristics of students correlate with the ability composition of teams. Again, there are no systematic differences between teams of different ability composition. GPA is a logical exception. Since the randomization checks in this subsection fail to find any pretreatment differences (that may contaminate the design of our field experiment), the analyses in the next section do not include control variables (adding superfluous controls would only reduce the degrees of freedom).

**Table 3.** Randomization checks at the individual and team level

	CV ability		CV ability <sup>2</sup>		Average ability	
<b>A1: Low ability students</b>						
Age	-4.825	(5.894)	16.482	(10.468)	-0.029	(0.068)
Gender	1.591	(1.293)	-4.452*	(2.394)	0.029*	(0.016)
Risk aversion	-3.301	(6.071)	13.287	(12.401)	-0.048	(0.070)
Grade point average	0.227	(0.807)	-1.311	(1.462)	-0.003	(0.010)
<b>A2: High ability students</b>						
Age	7.163	(5.515)	-10.984	(10.448)	-0.057	(0.049)
Gender	-1.682	(1.992)	3.388	(4.653)	-0.020	(0.017)
Risk aversion	-1.639	(6.008)	5.894	(11.588)	0.009	(0.070)
Grade point average	0.755	(0.499)	-0.874	(0.991)	0.005	(0.006)
<b>B: Team level (average)</b>						
Age	-0.048	(5.651)	5.631	(11.485)	-0.009	(0.047)
Gender	-0.537	(2.333)	0.517	(5.366)	-0.009	(0.012)
Risk aversion	-3.195	(6.361)	10.764	(12.887)	-0.009	(0.048)
Grade point average	0.371	(0.593)	-0.789	(1.280)	0.020***	(0.005)
Team size	-1.531	(18.318)	8.077	(39.221)	-0.057	(0.135)

*Note:* Average and CV of ability reflect at the team level respectively average score and coefficient of variation in scores on Raven’s advanced progressive matrices test. In panels A1 and A2 the coefficients come from a regression at the individual level of the row variable on the column variables, separately for students of low (test score  $\leq 18.60$ ) and high (test score  $> 18.60$ ) cognitive ability (robust standard errors in parentheses). In panel B, the coefficients come from a regression at the team level of the row variable on the column variables (bootstrapped standard errors in parentheses; 1000 replications). \*\*\*/\*\*/\* denotes significance at the 1%/5%/10%-level.

### 3 Results

#### 3.1 Main findings

Table 4 reports regression results for the effect of ability dispersion on the four measures of business performance that we study: sales, profits, a binary indicator for positive profits and profits per share. Panel A presents the results from regressing these performance measures on teams’ average test score, their coefficient of variation in test scores and its square. Besides standard OLS regression, we employ median and robust (M-estimation) regression to assess whether the results are sensitive to outliers.

Column (1) shows that sales first increase with ability dispersion up to a coefficient of variation in test scores of approximately 0.25 (the sample average is 0.22) and then decrease with ability dispersion, given teams’ average ability. Teams’ average ability levels show consistently positive but

insignificant effects in these specifications. However, columns (2) and (3) indicate that this effect of ability dispersion on sales tends to be inflated by outliers: the point estimates are insignificant when using median and robust (M-estimation) regression estimation techniques. Columns (4) through (6) consistently show an inverse U-shaped pattern for the relationship between ability dispersion and profits. Again, performance is maximized at a coefficient of variation in test scores of about 0.25. The same holds for the probability of profits being positive in column (7), although the degree of ability dispersion where performance peaks marginally increases to a coefficient of variation in test scores of 0.27. The coefficients in columns (8) through (10) corroborate these findings: the effect of ability dispersion on profits per share is described by an inverse U-shape with the optimum at a coefficient of variation in test scores of roughly 0.25. The results from the quadratic specifications in columns (4) through (10) are robust to outliers. Similar results are obtained when we exclude teams' average ability or include higher-order terms for the average ability of teams (not tabulated).

One limitation of a quadratic specification is that it forces symmetric effects of ability dispersion below and above the estimated maximum. In panel B we estimate spline functions allowing these effects of ability dispersion below (1st segment) and above (2nd segment) the maximum (a coefficient of variation in test scores of 0.25) to be increasing and decreasing at different rates. The cutoff in our spline functions is obtained by averaging the coefficients of variation in cognitive ability that maximize team performance (according to the quadratic specifications). Results from these spline functions indicate that business performance tends to increase with ability dispersion below a coefficient of variation in test scores of 0.25. If the coefficient of variation in test scores is at least equal to 0.25 all coefficients for the impact of ability dispersion are negative and (with three exceptions) significant. The point estimates in column (5) of panel B imply that raising the coefficient of variation in test scores from 0.20 to 0.25 increases profits by about 200 euros (approximately half of a standard deviation), while profits decrease by roughly the same amount if the coefficient of variation in test scores is further raised from 0.25 to 0.30. In sum, the results of panel B closely resemble the inverse U-shaped pattern from the quadratic specifications of panel A, although the limited number of teams may slightly reduce the precision of its estimates.



**Table 4.** Ability dispersion and team performance

	Sales			Profits			Pos. profits	Profits per share		
	OLS (1)	Median (2)	Robust (3)	OLS (4)	Median (5)	Robust (6)	OLS (7)	OLS (8)	Median (9)	Robust (10)
<b>A: Polynomial</b>										
CV ability	14772.5* (8083.3)	4439.5 (5549.2)	4422.5 (5217.3)	8243.7** (3543.9)	7378.1** (3200.0)	6052.4* (3233.6)	8.952*** (3.163)	106.100* (58.466)	92.478* (52.612)	89.344 (57.936)
CV ability <sup>2</sup>	-30305.8* (17055.2)	-9485.4 (10975.2)	-9118.5 (11357.6)	-16025.3** (6795.2)	-14878.5** (6385.5)	-12441.4** (6276.4)	-16.677** (6.580)	-207.870* (112.727)	-196.543* (104.931)	-178.400 (115.048)
Avg. ability	25.529 (62.982)	3.449 (50.458)	-2.862 (44.774)	19.528 (25.975)	18.594 (26.966)	4.261 (25.490)	0.004 (0.032)	0.252 (0.461)	0.105 (0.475)	0.115 (0.467)
<i>Maximum</i>	<i>0.24</i>	<i>0.23</i>	<i>0.24</i>	<i>0.26</i>	<i>0.25</i>	<i>0.24</i>	<i>0.27</i>	<i>0.26</i>	<i>0.24</i>	<i>0.25</i>
<i>R</i> <sup>2</sup>	<i>0.09</i>	<i>0.04</i>	<i>0.03</i>	<i>0.15</i>	<i>0.08</i>	<i>0.11</i>	<i>0.14</i>	<i>0.09</i>	<i>0.08</i>	<i>0.07</i>
<b>B: Spline (0.25)</b>										
1st segment	7377.2* (3788.4)	2629.7 (2486.5)	1976.8 (2310.0)	3817.7** (1895.2)	3845.9** (1583.8)	2695.7 (1806.3)	3.892*** (1.508)	51.036* (30.705)	46.191 (28.296)	50.538 (30.742)
2nd segment	-8317.5** (4060.6)	-2818.4 (3323.5)	-2247.2 (3133.5)	-3352.9** (1694.5)	-4022.5** (1560.1)	-2879.5* (1724.4)	-2.576 (1.796)	-47.116* (28.036)	-59.334** (25.836)	-46.582* (28.026)
Avg. ability	42.106 (67.387)	29.512 (51.983)	0.560 (45.011)	25.513 (27.893)	36.167 (29.403)	12.321 (27.956)	0.007 (0.034)	0.342 (0.515)	0.209 (0.516)	0.349 (0.543)
<i>R</i> <sup>2</sup>	<i>0.14</i>	<i>0.05</i>	<i>0.04</i>	<i>0.17</i>	<i>0.09</i>	<i>0.13</i>	<i>0.13</i>	<i>0.11</i>	<i>0.10</i>	<i>0.12</i>

*Note:* Based on information from 49 teams. Average and CV of ability reflect at the team level respectively average score and coefficient of variation in scores on Raven's advanced progressive matrices test. OLS, Median and Robust refer to the estimation method. Median and robust specifications for positive profits are excluded since this variable is dichotomous. Bootstrapped standard errors in parentheses (1000 replications). \*\*\*/\*\*/\* denotes significance at the 1%/5%/10%-level.

### 3.2 Robustness

Our first set of robustness checks explores other measures of ability dispersion, including teams' standard deviation of ability (Hansen et al., 2006); teams' ratio of the maximum to the minimum ability (Hamilton et al., 2003); and spline functions with three segments of ability dispersion (cutoff levels at a coefficient of variation in test scores of 0.15 and 0.30). Second, we study whether there is a relationship between the ability dispersion within teams and the occurrence of dismissals of individual team members. To the extent that team members are productive, dismissals may be viewed as an indirect and inverse measure of team performance. Therefore one would expect, if anything, that ability diversity has a U-shaped relationship with dismissals.

Panels A and B of Table A2 in the appendix reveal an inverse U-shaped effect of teams' standard deviation in ability and teams' ratio of the maximum to the minimum ability on their performance, although significance levels vary across both measures of ability dispersion.<sup>9</sup> The degree of ability dispersion that maximizes team performance is again slightly above the sample average. Panel C indicates a similar inverse U-shaped pattern for spline functions with teams of low, moderate and high ability dispersion (based on the coefficient of variation in test scores). The point estimates for teams of low ability dispersion ( $CV < 0.15$ ) are positive and relatively large compared to the coefficients for teams of moderate ability dispersion ( $0.15 \geq CV \leq 0.30$ ). For teams of high ability dispersion ( $CV > 0.30$ ) the point estimates are negative and also relatively large in relation to those for teams of moderate ability dispersion. The number of teams, however, limits the precision of these estimates. In sum, although significance levels are consequently muted for some of these dispersion measures, none of these robustness checks are at odds with the results previously obtained.

Panel A of Appendix Table A3 shows results of estimating the relationship between ability dispersion and dismissals. The number and incidence of dismissals reflect respectively the number of dismissals per team and whether or not a team has experienced at least one dismissal (dummy = 1 if any). Panel A shows for both number and incidence that teams of moderate ability dispersion are characterized by fewer dismissals and, we infer, less free-riding. Dismissals do not reflect a process whereby teams simply get rid of low ability or high ability students (rather than shirkers)

---

<sup>9</sup>The ratio of the maximum to the minimum ability in the team is the most sensitive to outliers since this measure of ability dispersion could already be considerably inflated by only one team member of (very) low or high cognitive ability.

since the relationship between cognitive ability and probability of dismissal at the individual level is insignificant (not tabulated). Hence, dismissals do not change the composition of teams in terms of their cognitive abilities. The number and incidence of dismissals are minimized at a coefficient of variation in test scores of approximately 0.24. Note that this minimum almost exactly corresponds with the coefficient of variation in test scores that maximizes business performance (about 0.25). This result is consistent with the main results in Table 4 if dismissals are an indirect and inverse measure of business performance. Panel B indicates that, indeed, fewer dismissals are associated with better business performance, separately for number and incidence. However, we lack exogenous variation to identify a causal impact of dismissals on business performance since dismissals are obviously endogenous.

## 4 Theory

This section proposes a simple model intended to understand how business team performance varies with intra-team ability dispersion. We deliberately study a setting which is similar to that of our field experiment in several important respects.

### 4.1 Set-up

A team (or firm) employs  $n \geq 2$  workers of potentially different abilities. To start with, and following both previous work (e.g., Prat, 2002) and our field experiment,  $n$  will be taken as fixed and exogenous. Below, in section 4.3,  $n$  will be endogenized. Teams allocate members among several distinct tasks indexed by  $i$ . As in Hong and Page (2004), each person performs only one task. The set of members performing the same task  $i$  is referred to as a ‘sub team’: the cardinality of this set is  $m_i$ , where  $m_i < n$ . Sub-team  $i$  contributes output  $X_i$  to the team. Total team value  $V$  is the sum of these outputs, i.e.,  $V = \sum_i X_i$ . This captures the idea that output in business settings typically involves a combination of several different activities conducted by different people.<sup>10</sup>

Tasks differ from each other in terms of their inherent productive values. Let task 1 be the most valuable task, task 2 the next most valuable, and so on. A person working on task  $i$  generates an

---

<sup>10</sup>We deliberately study an additive structure thereby removing a well-known rationale for diverse teams based on complementarities between different tasks (Prat, 2002). Note also that allowing for unequal weights of tasks in  $V$  would not affect our central results.

outcome for that task which is a draw from an exponential distribution with parameter

$$\lambda_i = 1 + \epsilon(i - 1),$$

where  $\epsilon > 0$  can be regarded as a measure of productivity differences between tasks. The higher is  $\epsilon$  the more dispersed is productivity across tasks. The exponential distribution is convenient because the expected value of this distribution with parameter  $\lambda_i$  is simply  $1/\lambda_i$ , which will simplify the exposition below.

Analogous to Hong and Page (2004), each task output can be regarded as the best solution to a (stochastic) search problem undertaken independently by each of  $m_i < n$  members assigned to it. That is, the output of any task is the maximum draw from the outputs of each member who works on it. This has a straightforward interpretation: there are usually several ways a task can be executed, and sub-teams are interested in selecting the best out of those proposed by their members. We will initially assume that, regardless of their ability, each person can be expected to perform any task equally well as another, and so can be assigned to any task among those available. This assumption is consistent with the field experiment in which team members are not assigned to tasks on the basis of their abilities, and can be rotated between management and non-management tasks.<sup>11</sup>

Given this set-up, some restriction on the formation of sub-teams is necessary to keep the problem non-trivial, since if anyone can work in any team, all teams can perform equally well regardless of their diversity. We propose a restriction noting that in an entrepreneurial setting such as that of our field experiment, team members often have some discretion about the other team members they work closely with on a particular task. In practice, people often partner with people like themselves (Ruef et al., 2003), a tendency known as ‘assortative matching’. For example, evidence shows that people of a given cognitive ability tend to associate with others of similar cognitive ability (Lou et al., 1996). Reasons include performance benefits of assortative matching, even among people in exclusively low-ability groups; enhanced trust among cognitively similar people; and greater ease of communication between them (Ibarra, 1992; Lou et al., 1996).

---

<sup>11</sup>The rotation feature of the experiment does not affect the predictions of the model as long as abilities do not directly affect task outputs, as we have assumed. Nevertheless, we will relax this assumption in section 4.3, by allowing individual ability differences also to affect task productivity directly.

If assortative matching were perfect, teams would not face a decision problem at all, because sub-teams would be perfectly homogeneous. There would be a (given number)  $m_i$  people engaged in sub-task  $i$ , and teams would be defined by the set  $\{m_i : i = 1, \dots, d\}$ , where  $d$  is the number of different ability levels in a team, and  $n = \sum_{i=1}^d m_i$ . This set is given exogenously in our field experiment; hence there would be no member allocation problem to solve. And, as noted above, if at the other extreme there were no assortative matching at all, diversity would be irrelevant as members could be allocated freely over all tasks, and all teams would optimally choose the same allocation regardless of their diversity.

To generate an interesting problem, we therefore assume *imperfect assortative matching*, which takes the following form. Suppose there are  $d$  different ability types where  $d$ , fixed exogenously, is a natural measure of team diversity. Imperfect assortative matching is taken to entail an aversion by some individuals to working alongside other ability types. Specifically, we assume that at least one member of every ability group strictly prefers to work in a sub-team associated with their own ability type — even if that means working alone. Hence, a team with diversity  $d$  comprises exactly  $d$  sub-teams. However, matching is not perfect, in the sense that *some* members are willing to switch into different sub-teams among the  $d$ .

An example might help at this juncture. Let  $n = 5$  and denote abilities by  $\mathbf{a} = (a^1, a^1, a^2, a^2, a^3)$ , where superscripts denote different ability groups. If there were perfect assortative matching, the only possible configuration would be  $(m_1, m_2, m_3) = (2, 2, 1)$ . But with imperfect assortative matching, configurations like  $(3, 1, 1)$  and  $(1, 2, 2)$  are also possible.

To summarize, each team faces a different distribution of abilities but a common decision problem, i.e., how to assign members to sub-tasks subject to: a given set of abilities of its members ( $n$  and  $d$  both exogenous); imperfect assortative matching giving rise to  $d$  sub-teams; and a set of exogenous task productivities parameterized by  $\epsilon$ . We analyze this decision problem next.

#### 4.2 Team performance

We now derive expressions for expected team performance. Consider first a single sub-team  $i$ : each member  $j$  working in this sub-team takes an independent draw from the same exponential distribution with parameter  $\lambda_i$ , generating outcomes  $X_{ij}$ :  $j = 1, \dots, m_i$ . The expected value of the

sub-team is  $E(X_i)$ , where we define

$$X_i = \max \{X_{i1}, X_{i2}, \dots, X_{im_i}\}$$

Letting  $F_{X_i}(x)$  denote the distribution function of  $X_i$ , we have

$$\begin{aligned} E(X_i) &= \int_0^\infty [1 - F_{X_i}(x)] dx \\ &= \int_0^\infty \left[ 1 - \prod_{j=1}^{m_i} \Pr(X_{ij} < x) \right] dx \\ &= \int_0^\infty \left[ 1 - (1 - e^{-\lambda_i x})^{m_i} \right] dx \\ &= \frac{1}{\lambda_i} \sum_{j=1}^{m_i} \frac{1}{j}. \end{aligned} \tag{1}$$

The last line is non-trivial and uses properties of generating functions: see, e.g., Lugo (2011) for a proof.

We can now state the expected value of the entire team,  $E(V)$ . By the independence and additivity of the  $X_i$ , and using (1), teams solve the program:

$$\begin{aligned} \max_{\{m_i: i=1, \dots, d\}} E(V; d) &= \max_{\{m_i: i=1, \dots, d\}} \left\{ \sum_{i=1}^d \frac{1}{\lambda_i} \sum_{j=1}^{m_i} \frac{1}{j} \right\} \\ \text{s.t.} \quad &0 \leq m_i \leq n : i = 1, \dots, d; d \geq 1, n \geq 2 \text{ given} \\ &\sum_{i=1}^d m_i = n. \end{aligned} \tag{2}$$

At the heart of (2) lies a trade-off between specialization and diversification. On the one hand, there are benefits from having large specialized sub-teams of workers who focus on the most valuable task. On the other hand this strategy eventually encounters diminishing returns and ‘crowds out’ (given fixed  $n$ ) other people who perform less valuable yet still positive productivity tasks in different sub-teams, who have not yet run into diminishing returns. But, benefits of diversification are limited by the rate at which task productivity declines as additional tasks are performed.

The basic logic behind the model can be seen even in the simplest case of  $n = 2$ . The configuration  $\{m_1, m_2\} = \{2, 0\}$  is a team comprised of two ability types, both of whom work on the

most productive task,  $i = 1$ . For this team,  $d = \lambda_1 = 1$  and  $E(V) = 1 + 0.5 = 1.5$ . The alternative  $d = 1$  configuration  $\{0, 2\}$  on the other hand yields  $E(V) = 1.5/(1 + \epsilon) < 1.5$ . Finally, consider a different ability distribution with  $\{1, 1\}$ , the most diverse of all two-person teams,  $d = 2$ : this yields  $E(V) = 1 + [1/(1 + \epsilon)]$ . Whether the  $\{1, 1\}$  configuration is preferable to the  $\{2, 0\}$  configuration evidently depends on the value of  $\epsilon$ . If  $\epsilon < 1$ , the diverse team  $\{1, 1\}$  yields higher performance; but if  $\epsilon > 1$ , the specialized team  $\{2, 0\}$  is preferred.

This example illustrates a point which remains true for all  $n > 2$ : the lower  $\epsilon$  is, i.e., the more similar tasks are with respect to their productivities, the more slowly payoffs to team diversity decrease, relative to the marginal rate of specialization payoffs (which decreases at rate  $1/m_i$ ). Conversely, the higher  $\epsilon$  is, the faster payoffs to diversity decrease relative to those under specialization. Hence, in general we might expect to see: specialized teams emerge in environments where task productivities are very different; diverse teams in environments where task productivities are very similar; and moderately diverse teams in environments where differences between task productivities are moderate.

Column (1) of Table 5 illustrates solutions to (2) for the case of  $n = 5$  and various values of  $\epsilon$ . Panel A of Table 5 reports results for  $\epsilon = 0.1$ , corresponding to small differences in productivity between tasks. Panel B reports results for moderate differences and Panel C large differences between tasks. For each panel, column (1) reports optimal team configurations,  $m^*$ , and maximal  $E(V)$  for each diversity level  $d$  — from  $d = 1$  (maximum specialization) through  $d = 5$  (maximum diversity). The results in column (1), Panel B show that an intermediate degree of diversity ( $d = 4$ ) maximizes expected performance with a value of 3.07. As noted above, this is exactly what would be expected when differences in task productivity are moderate; and it is consistent with the findings from our field experiment. In panel A, differences in task productivity are very small, so maximal diversity is strictly preferred. Differences in task productivity are very large in the third panel, where in accordance with the logic above minimal diversity becomes optimal.

#### 4.3 Extensions of the model

Having elucidated the conditions under which the model generates findings consistent with our field experiment, we next extend it in two ways. First, we allow individual differences in ability also to affect the productivity of each task directly. As above, let  $\mathbf{a}$  denote the team's  $n$ -vector

**Table 5.** Team diversity and expected performance

	(1)		(2)			(3)		
	$m^*$	$E(V; d, \gamma = 0)$	$E(V; d, \gamma = 0.5)$	$E(V; d, \gamma = 1)$	$E(V; d, \gamma = 2)$	$m^*$	$n^*$	$E(V; d, w = 0.38)$
A. $\epsilon = 0.1$								
$d = 1$	(5, 0, 0, 0, 0)	2.28	2.26	2.27	2.41	(2, 0, 0, 0, 0)	2	0.74
$d = 2$	(3, 2, 0, 0, 0)	3.20	3.17	3.20	3.38	(2, 2, 0, 0, 0)	4	1.34
$d = 3$	(2, 2, 1, 0, 0)	3.70	3.67	3.69	3.91	(2, 2, 2, 0, 0)	6	1.83
$d = 4$	(2, 1, 1, 1, 0)	4.10	3.99	4.00	4.25	(2, 2, 2, 2, 0)	8	2.22
$d = 5$	(1, 1, 1, 1, 1)	4.22	4.20	4.23	4.46	(2, 2, 2, 2, 1)	9	2.55
B. $\epsilon = 0.5$								
$d = 1$	(5, 0, 0, 0, 0)	2.28	2.26	2.27	2.41	(2, 0, 0, 0, 0)	2	0.74
$d = 2$	(3, 2, 0, 0, 0)	2.83	2.81	2.83	2.99	(2, 1, 0, 0, 0)	3	1.03
$d = 3$	(3, 1, 1, 0, 0)				3.16	(2, 1, 1, 0, 0)	4	1.15
	(2, 2, 1, 0, 0)	3.00	2.98	2.99				
$d = 4$	(2, 1, 1, 1, 0)	3.07	3.04	3.07	3.23	(2, 1, 1, 1, 0)	5	1.17
$d = 5$	(1, 1, 1, 1, 1)	2.90	2.88	2.89	3.08	(2, 1, 1, 1, 1)	6	1.12
C. $\epsilon = 5$								
$d = 1$	(5, 0, 0, 0, 0)	2.28	2.26	2.27	2.41	(2, 0, 0, 0, 0)	2	0.74
$d = 2$	(4, 1, 0, 0, 0)	2.25	2.23	2.25	2.38	(2, 1, 0, 0, 0)	3	0.53
$d = 3$	(3, 1, 1, 0, 0)	2.09	2.07	2.09	2.21	(2, 1, 1, 0, 0)	4	0.24
$d = 4$	(2, 1, 1, 1, 0)	1.82	1.81	1.84	1.92	(2, 1, 1, 1, 0)	5	-0.08
$d = 5$	(1, 1, 1, 1, 1)	1.32	1.36	1.37	1.46	(2, 1, 1, 1, 1)	6	-0.41

*Note:* Columns (1) and (2) are for exogenous  $n = 5$ : optimal  $m^*$  values in column (1) apply to column (2) as well, except for  $d = 3$  in Panel B, where only one of the two  $m^*$  vectors is optimal for any given  $\gamma$ . Column (3) is for endogenous  $n$ ; different  $m^*$  solve this problem, given in the first sub-column of (3).

of abilities drawn at random from the population: denote mean ability by  $\mu$ . Let  $a_i \in \mathbf{a}$  be the ability of a member assigned to task  $i$ . There are various ways that individual abilities could influence productivity; we choose the following simple specification which makes productivity a strictly increasing function of ability:

$$\max_{\{m_i: i=1, \dots, d\}} E(V; d, \gamma) = \max_{\{m_i: i=1, \dots, d\}} \left\{ \sum_{i=1}^d \frac{(a_i/\mu)^\gamma}{\lambda_i} \sum_{j=1}^{m_i} \frac{1}{j} \right\}, \quad (3)$$

where  $\gamma > 0$  is a shape parameter. Note that (3) collapses into (2) when  $\gamma = 0$ .

How do the results change when (3) replaces (2) for  $\gamma \neq 0$ ? To explore this, we draw  $n$ -vectors  $\mathbf{a}$  using data from the field experiment. Hence, for the illustrative case of  $n = 5$ , we use quintiles of the actual ability distribution and set  $\mu = 18.6$ . Noting that it is optimal to assign the ablest people to the most productive tasks, column (2) of Table 5 reports the results of solving the programming



problem using (3) and several values of  $\gamma$ . As can be seen, the same patterns are observed as in column (1): moderate task productivity differences (Panel B) once again yield an inverse-U shape relationship between expected performance and team diversity, while that is not the case if task differences are small (Panel A) or large (Panel C). Hence, the predictions of the model appear to be robust to allowing ability to have direct as well as indirect effects on team production.

Turning to the second extension of the model, we note that although our field experiment embodies some salient aspects of real-world teams, one exception is the experiment’s restriction of a fixed number of team members,  $n$ . We next extend the model to allow  $n$  to be endogenous, in a setting where team members all cost the same exogenous wage  $w$ . Then the objective in (2) changes to become

$$\max_{\{m_i:i=1,\dots,d\},n} \left\{ \sum_{i=1}^d \left( \frac{1}{\lambda_i} \sum_{j=1}^{m_i} \frac{1}{j} - w m_i \right) \right\} \quad (4)$$

Essentially, the team optimally hires additional members until the marginal product of an additional member in task  $i$  equals  $w$ . Although we do not observe  $w$  for this sample of students, we can choose a value which generates a non-trivial problem. Specifically, note that  $w \rightarrow 0$  returns us to (2), while if  $w \rightarrow \infty$ ,  $m_i \rightarrow 0 \forall i$ . An intermediate value of  $w = 0.38$  is of comparable magnitude to the other elements of (4) and leads to non-trivial solutions, reported in column (3) of Table 5. As can be seen there, although team sizes vary quite a lot, the same qualitative results emerge again. In particular, with an intermediate degree of task productivity diversity (Panel B), we once again observe an inverse-U shaped relationship between expected team performance and team diversity, with the same optimal solution of  $d = 4$ . Results for the other panels of column (3) with endogenous  $n$  are similar to the corresponding panels of columns (1) and (2).

To summarize, we have shown that the model’s predictions are both consistent with the findings of our field experiment and robust to two natural and realistic extensions to the model. Of course, other extensions could doubtless be proposed; we discuss some of these in our closing discussion.

## 5 Discussion and conclusion

This paper presented the results of a field experiment in which 573 students in 49 teams started up and managed real companies under identical circumstances. We ensured exogenous variation in — otherwise random — team composition by assigning students to teams based on their measured

cognitive abilities (Raven test). Our experiment likely measures the causal effect of ability dispersion on team performance in a setting that closely resembles the functioning of business teams which simultaneously undertake a range of diverse tasks. Our key finding was that team performance exhibits an inverse U-shaped relationship with ability dispersion; average cognitive ability of team members did not significantly improve performance. Performance was maximized with a coefficient of variation in cognitive ability of about 0.25 — close to the sample average of 0.22.

In contrast to the predictions from prior theoretical work, we did not find empirical support for the dominance either of maximally homogeneous teams (Prat, 2002) or maximally diverse teams (Hong and Page, 2004). In an effort to better understand the processes at work and understand and possibly reconcile the differences between our and earlier work, we developed a simple model in which teams perform several tasks of varying productivity. Incorporating into the model the well-known tendency of people to sort into similar groups (Ruef et al., 2003), we proposed that our empirical findings are most likely to emerge when productivity differences between disparate tasks are moderate. Different predictions, associated with monotonic relationships between team diversity and performance, only emerge when productivity differences between tasks are either very small or very large.

We believe this paper carries several implications for both scholarship and practice. To start with, the use of a field experiment with stratified randomization is rather new for the empirical analysis of team performance. We hope that future empirical research will emulate the benefits of such an approach, especially its ability to overcome self-selection problems which have bedeviled more traditional empirical approaches. Second, our findings challenge the tendency of prior empirical studies to look for simple monotonic relationships between team diversity and performance outcomes. In a sense, to ask whether team diversity is good or bad for performance might be to pose the wrong question: instead, our results suggest that it may be more sensible to ask how much diversity is conducive to team performance given circumstances such as task diversity. Indeed, this insight finds a parallel in the theoretical model we proposed to understand the findings. Rather than suggesting that diversity — even moderate levels of it — is invariably beneficial, the model instead predicts environment-specific outcomes. That is, our model comes with a boundary condition which determines the qualitative nature of the relationship between team diversity and performance.

Third, our empirical findings may also be of interest to practitioners: entrepreneurial teams in practice are usually fairly homogeneous in nature (Ruef et al., 2003), though (in accordance with the assumptions of the model) far from perfectly so. Unlike prior work, which usually counsels practitioners that entrepreneurial teams risk being insufficiently diverse, our work anticipates the possibility that such teams might actually be configured with an appropriate degree of diversity. In addition, our model suggests that the concept of diversity can also be understood in terms of the allocation of members among tasks. It may therefore be fruitful for practitioners to benchmark ‘balanced’ teams not only in terms of the distribution of characteristics of their members, but also in terms of the number and diversity of tasks they perform and the allocation of member effort among those tasks.

A similar implication carries over into managerial settings where managers select team members but where those members have some discretion over whom they partner with in sub-teams. In such cases, the imperfect assortative matching model we have postulated may be pertinent, especially if there is asymmetric information which prevents managers from perfectly monitoring and altering sub-team composition. In such cases, managers might do best by working with the grain of imperfect assortative matching, by selecting strategically members into teams in order to generate a desired degree of diversity. As our model has shown, the desired degree of diversity depends on the production structure in a multi-task environment — a structure of which managers are presumably cognizant.

There are of course several limitations of our study. To start with, the model assumed a particular form of imperfect assortative matching; predictions may differ under alternative forms of this assumption. The model also abstracted from multi-tasking by individuals. While this was not important in the ‘vanilla flavor’ of our model where there were no direct effects of ability on production, future conceptual work can explore how multi-tasking affects optimal team composition when abilities do directly influence productivity in different tasks.

Second, the field experiment did not generate data about the tasks that team members performed in practice, and exactly how team members were assigned to them in practice, including the manifestation of imperfect assortative matching. Future research, possibly in a laboratory setting, is needed to delve deeper into these deep aspects of team production.

Third, the external validity of our empirical findings is another limitation of the study. We

exploited exogenous variation in cognitive ability among students who are young, and generally lack work experience. The sorting behaviors and performance outcomes of these agents might not be representative of teams comprised of older, more experienced people. For this reason, future research might seek to replicate experiments like ours in real organizations and on a larger scale. Yet, it should also be borne in mind that the subjects of our experiment did engage in a substantial business project requiring execution of a broad array of non-trivial tasks; and they also faced strong incentives as would generally apply in broader real-world settings too. Finally, one might argue that our results are not generalizable because the student population that we study might be representative of only a limited part (the higher end) of the ability distribution that one faces usually in the workplace. However, based on statistics about the student population of first-year students in a vocational college (such as we study), their entry levels, attrition rates, and the likelihood of moving on to higher degrees, we found that the education distribution has a slightly lower variance and higher mean than the general population (including those adults not active in the labor force). So all in all, and despite these limitations, we believe that our field experiment remains informative about the impact of ability dispersion on the performance of business (management) teams.

There are several other fruitful areas for future research into team composition and performance. It would be interesting, for example, to analyze interactions between team members, both theoretically and empirically. Possible interactions include knowledge transfers, spillovers, conflicts and communication problems. Our model abstracted from these issues. While post-experiment interview transcripts did not reveal a dominant role in practice for any of them, it is possible that they might be more salient in other settings. Furthermore, experimental work which varies productivity differences between tasks could check whether the relationship between team diversity and performance does indeed switch qualitatively, as predicted by our model.

To conclude, a whole array of future questions about team dynamics and performance remains to be explored, including the role of tastes and preferences, beliefs and (re)negotiation. These promise to deepen further our understanding of how teams work and the best way that teams can be configured in different circumstances to maximize performance. Scholars, entrepreneurs and managers all stand to gain from the further development of this research agenda.

## References

- Bors, D. and Stokes, T. (1998). Raven's advanced progressive matrices: Norms for first-year university students and the development of a short form. *Educational and Psychological Measurement*, 58(3):382–398.
- Calvin, C., Deary, I., Fenton, C., Roberts, B., Der, G., Leckenby, N., and Batty, G. (2011). Intelligence in youth and all-cause-mortality: Systematic review with meta-analysis. *International Journal of Epidemiology*, 40(3):626–644.
- Carpenter, P., Just, M., and Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven progressive matrices test. *Psychological Review*, 97(3):404–431.
- Cattell, R. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54(1):1–22.
- Franck, E. and Nüesch, S. (2010). The effect of talent disparity on team productivity in soccer. *Journal of Economic Psychology*, 31(2):218–229.
- Gould, E. and Winter, E. (2009). Interactions between workers and the technology of production: Evidence from professional baseball. *Review of Economics and Statistics*, 91(1):188–200.
- Hamel, R. and Schmittmann, V. (2006). The 20-minute version as a predictor of the Raven advanced progressive matrices test. *Educational and Psychological Measurement*, 66(6):1039–1046.
- 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.
- Hamilton, B., Nickerson, J., and Owan, H. (2012). *Diversity and Productivity in Production Teams*, chapter in: A. Bryson (ed.), *Advances in the Economic Analyses of Participatory & Labor-Managed Firms*, volume 13, pages 99–138. Bingley: Emerald Group Publishing Limited.
- Hansen, Z., Owan, H., and Pan, J. (2006). The impact of group diversity on performance and knowledge spillover: An experiment in a college classroom. NBER Working Paper.

- HBO-raad (2010). Feiten en cijfers: Afgestudeerden en uitvallers in het hoger beroepsonderwijs. Technical report, HBO-raad.
- Hong, L. and Page, S. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 101(46):16385–16389.
- Hoogendoorn, S., Oosterbeek, H., and Van Praag, M. (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. and Van Praag, M. (2012). Ethnic diversity and team performance: A field experiment. IZA Discussion Paper.
- Ibarra, H. (1992). Homophily and differential returns: Sex differences in network structure and access in an advertising firm. *Administrative Sciences Quarterly*, 37(3):422–447.
- Irwing, P. and Lynn, R. (2005). Sex differences in means and variability on the progressive matrices in university students: A meta-analysis. *British Journal of Psychology*, 96(4):505–524.
- Kahn, L. (2000). The sports business as a labor market laboratory. *Journal of Economic Perspectives*, 14(3):75–94.
- Lou, Y., Abrami, P., Spence, J., Poulsen, C., Chambers, B., and Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Education Research*, 66(4):423–458.
- Lugo, M. (2011). The expectation of the maximum of exponentials. Course notes: [www.stat.berkeley.edu/~mlugo/stat134-f11/exponential-maximum.pdf](http://www.stat.berkeley.edu/~mlugo/stat134-f11/exponential-maximum.pdf), Retrieved: December 2013.
- Mills, C., Ablard, K., and Brody, L. (1993). The Raven’s progressive matrices: Its usefulness for identifying gifted/talented students. *Roeper Review*, 15(3):183–186.
- Nisbett, R., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D., and Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*, 67(2):1–30.

- 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.
- Papps, K., Bryson, A., and Gomez, R. (2011). Heterogeneous worker ability and team-based production: Evidence from major league baseball, 1920–2009. *Labour Economics*, 18(3):310–319.
- Prat, A. (2002). Should a team be homogeneous? *European Economic Review*, 46(7):1187–1207.
- Raven, J. (2000). The Raven’s progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, 41(1):1–48.
- Raven, J., Raven, J., and Court, J. (1993). *Manual for Raven’s Progressive Matrices and Vocabulary Scales: Section 1 General overview*. Oxford: Psychologists Press.
- Raven, J., Raven, J., and Court, J. (1998). *Manual for Raven’s Progressive Matrices and Vocabulary Scales: Section 4 Advanced Progressive Matrices*. Oxford: Oxford Psychologists Press.
- Ruef, M., Aldrich, H., and Carter, N. (2003). The structure of organizational founding teams: Homophily, strong ties and isolation among US entrepreneurs. *American Sociological Review*, 68(2):195–222.
- Rushton, J. and Jensen, A. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy and Law*, 11(2):235–294.
- Spearman, C. (1927). *The Abilities of Man*. New York: Macmillan.
- Woolley, A., Chabris, C., Pentland, A., Hashmi, N., and Malone, T. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330(6004):686–688.

**Appendix Table A1.** Team characteristics

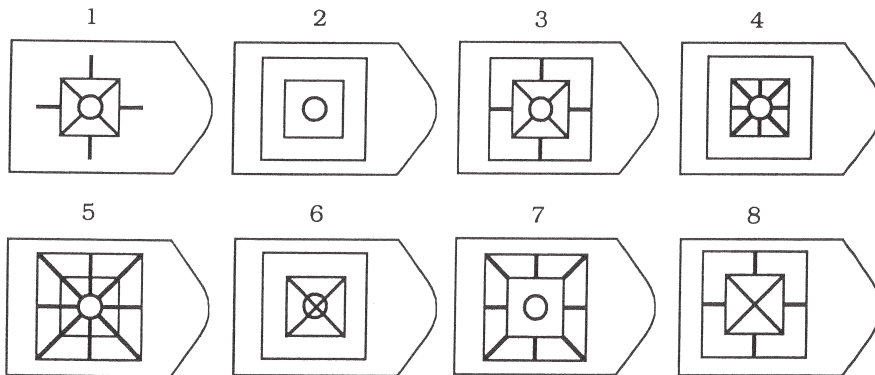
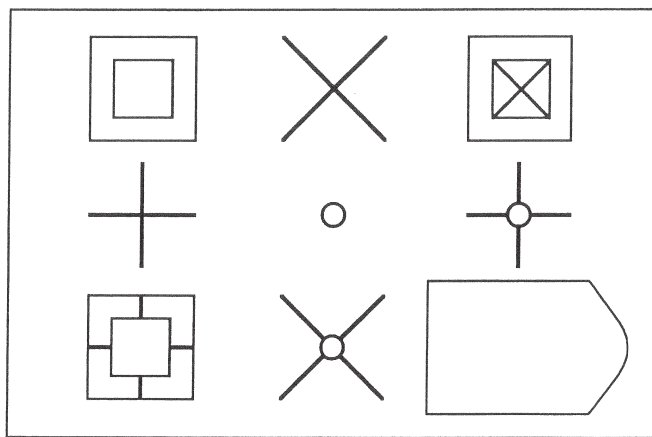
#	Name	Average ability	CV ability	Team size	Sales (euros)	Profits (euros)	Profits/share	Description of product/service
1	Achter de rits	17.64	0.33	11	533.95	78.20	1.56	Sustainable fair trade boxer short
2	Be Portable	16.00	0.22	10	1179.58	288.30	4.65	Portable durable water bottle
3	Berry Emotions	21.45	0.26	12	88.24	-591.58	-13.76	Personalized telephone cover
4	Blue Empire	14.09	0.19	13	418.07	-75.44	-1.80	Fashionable bluetooth bracelet
5	Cleaning Env. Friendly	14.75	0.31	8	415.76	3.72	0.06	Long-lasting durable paper towel
6	Chori-Chori	16.62	0.20	15	588.24	-441.00	-3.94	Trendy customized Indian jewelry
7	Comb. Young E'ship	15.13	0.19	9	894.12	119.90	2.45	Whistling operated key finder
8	Covers 'n Cases	19.36	0.37	16	338.87	56.10	1.52	Covers and cases for telephones
9	Easylife	19.13	0.32	12	207.18	-288.79	-8.25	Automated toothpaste dispenser
10	Flashing Productions	16.70	0.31	12	615.88	90.34	1.96	Fluorescent nightlife bracelet
11	FresHES	15.08	0.20	15	3505.35	145.51	3.10	Wine bottle shaped umbrella
12	Freshual	22.08	0.15	13	1599.66	352.81	4.58	Convenient compact key torch
13	Green	19.50	0.21	10	524.37	111.81	4.66	Unbreakable vase for flowers
14	Green Solutions	14.07	0.31	14	1576.38	375.65	9.89	Environmental awareness products
15	Horanje	17.38	0.33	14	463.87	36.55	0.52	Gift bag with Dutch souvenirs
16	Ibrella	17.18	0.18	12	238.66	-473.79	-6.97	Easy-to-wear mini umbrella
17	I-corporation	16.27	0.22	15	43.15	-636.81	-9.65	Convenient eye-drop dispenser
18	I'Magine	18.83	0.21	15	3267.66	702.89	5.81	Fashionable Brazilian flip-flops
19	Innovart	22.40	0.16	12	42.02	-8.40	-0.20	Broker/intermediary in art
20	Jikoni	18.56	0.20	10	861.05	54.26	1.36	Innovative kitchen appliances
21	Kamikare	17.50	0.18	10	577.98	27.05	0.52	Easy-to-use hair styling product
22	K'deau	19.83	0.32	17	878.00	120.00	1.74	Gift box with various gadgets
23	Keyen	19.00	0.13	13	1525.79	256.11	3.41	High-quality key finder device
24	Lightdancer	16.17	0.47	13	139.79	-595.77	-7.84	Jewelry with built-in laser beam
25	MagicTouch	21.86	0.08	7	539.50	-532.96	-5.49	Compact women-friendly vibrator



**Appendix Table A1.** (continued)

#	Name	Average ability	CV ability	Team size	Sales (euros)	Profits (euros)	Profits/share	Description of product/service
26	Muffle	17.78	0.36	11	1320.76	331.09	6.13	Trendy coffee to-go cup sleeve
27	Nailtastic	17.78	0.15	10	284.87	-284.63	-7.49	Nailpolish applicator help kit
28	New Young Pros	20.43	0.30	10	1476.48	794.00	12.41	Anti-slip mat for multiple devices
29	Razor2Go	18.42	0.08	13	387.68	-113.31	-1.83	Comfortable facial hair remover
30	ReWrite	14.75	0.25	11	1600.45	413.39	6.36	T-shirt with rewritable tekst area
31	Rising Sun	14.78	0.30	11	577.71	-113.11	-4.71	Mini high-definition pen camera
32	Sleeves	20.13	0.25	10	5048.50	1619.39	25.70	Public transport chip card holder
33	Stud. Offer Products	22.22	0.18	9	144.18	-217.74	-4.35	Aids awareness-raising key chain
34	Sunny & Co	14.00	0.25	10	653.19	194.73	4.23	Wristband with sunscreen notifier
35	SunVision	20.40	0.17	10	391.66	-197.47	-5.81	Affordable sunglasses on strength
36	Swops	23.22	0.12	9	557.15	112.69	1.28	Trendy digital slim sports watch
37	Cardholder Company	18.11	0.12	9	252.10	-140.00	-5.60	Cardholder wallet with mirror
38	The Mug	20.70	0.15	13	2019.27	508.46	5.98	Comfortable coffee to-go mug
39	TieSupply	20.38	0.07	10	433.74	-242.86	-1.80	Tie roll cylinder in a travel case
40	Toccare	21.13	0.12	15	463.85	-225.90	-5.65	Key chain with digital business card
41	Winterwarmth	20.11	0.21	13	893.28	480.50	14.56	Multi-purpose hot and cold gel pad
42	Wristbands4World	17.69	0.14	14	596.00	50.88	1.64	Environmental awareness wristband
43	Xihe	20.70	0.09	11	860.50	-993.02	-10.13	Innovative solar battery charger
44	XXL	21.08	0.19	12	2552.91	238.86	2.91	HES University hooded sweatshirt
45	BubbleBeat	17.75	0.35	8	0.00	-225.00	-3.88	n/a
46	Cotton&Class	21.25	0.30	10	1201.00	-352.00	-4.69	n/a
47	Decent	19.25	0.08	13	210.00	-2.00	-0.08	n/a
48	InOffice	21.11	0.15	10	475.00	154.00	7.33	n/a
49	WatchUs	21.75	0.10	13	716.67	192.50	8.02	n/a

*Note:* Average and CV of ability reflect at the team level respectively average score and coefficient of variation in scores on Raven's advanced progressive matrices test (excluding students whose test score is unknown). The number of students whose test score is unknown amounts to 60 (out of 573 students). Team size reflects the size of teams at baseline. Missing or incomplete descriptions of a team's product or service are indicated by 'n/a' (not available).



**Appendix Figure A1.** Example of a figure from Raven's advanced progressive matrices test

**Appendix Table A2.** Other measures of ability dispersion and team performance

	Sales			Profits			Pos. profits	Profits per share		
	OLS (1)	Median (2)	Robust (3)	OLS (4)	Median (5)	Robust (6)	OLS (7)	OLS (8)	Median (9)	Robust (10)
<b>A: Standard deviation</b>										
SD ability	955.5** (446.4)	370.4 (286.2)	239.7 (272.0)	548.5*** (202.6)	538.0** (206.4)	412.6** (169.9)	0.641*** (0.195)	7.529** (3.131)	6.324* (3.650)	6.010** (2.765)
SD ability <sup>2</sup>	-107.6** (48.4)	-43.7 (33.4)	-27.7 (32.3)	-59.4*** (22.0)	-65.1** (24.6)	-46.9** (19.2)	-0.068*** (0.024)	-0.827** (0.339)	-0.801* (0.424)	-0.679** (0.314)
Avg. ability	18.841 (57.584)	30.451 (44.307)	-6.546 (36.635)	11.389 (20.706)	-2.816 (23.714)	-0.248 (18.422)	-0.008 (0.027)	0.163 (0.386)	-0.083 (0.462)	0.013 (0.364)
<i>Max./Avg.</i>	<i>4.44/3.91</i>	<i>4.24/3.91</i>	<i>4.33/3.91</i>	<i>4.62/3.91</i>	<i>4.13/3.91</i>	<i>4.40/3.91</i>	<i>4.71/3.91</i>	<i>4.55/3.91</i>	<i>3.95/3.91</i>	<i>4.43/3.91</i>
<i>R</i> <sup>2</sup>	<i>0.08</i>	<i>0.03</i>	<i>0.02</i>	<i>0.15</i>	<i>0.06</i>	<i>0.12</i>	<i>0.16</i>	<i>0.10</i>	<i>0.07</i>	<i>0.08</i>
<b>B: Ratio maximum/minimum</b>										
Ratio ability	2208.6** (1043.0)	1037.6 (787.2)	709.8 (788.1)	843.1 (550.5)	386.5 (571.2)	512.0 (606.1)	0.849 (0.683)	10.392 (9.255)	7.459 (10.154)	7.608 (10.879)
Ratio ability <sup>2</sup>	-403.5* (212.2)	-181.8 (164.0)	-131.9 (167.3)	-151.5 (114.9)	-58.3 (124.1)	-91.0 (130.6)	-0.140 (0.146)	-1.851 (1.908)	-1.127 (2.214)	-1.342 (2.352)
Avg. ability	43.312 (56.444)	40.983 (41.184)	7.209 (41.247)	18.076 (23.082)	13.565 (26.870)	3.956 (22.130)	0.003 (0.033)	0.242 (0.438)	0.389 (0.530)	0.089 (0.436)
<i>Max./Avg.</i>	<i>2.74/2.17</i>	<i>2.85/2.17</i>	<i>2.69/2.17</i>	<i>2.78/2.17</i>	<i>3.31/2.17</i>	<i>2.81/2.17</i>	<i>3.03/2.17</i>	<i>2.81/2.17</i>	<i>3.31/2.17</i>	<i>2.83/2.17</i>
<i>R</i> <sup>2</sup>	<i>0.10</i>	<i>0.05</i>	<i>0.04</i>	<i>0.08</i>	<i>0.03</i>	<i>0.04</i>	<i>0.08</i>	<i>0.04</i>	<i>0.03</i>	<i>0.03</i>
<b>C: Spline (0.15 and 0.30)</b>										
1st segment (N=12)	9022.6* (4649.3)	2193.0 (4643.2)	186.9 (4629.7)	5166.3* (2702.6)	5085.9 (3112.3)	3918.4 (2754.7)	7.297** (2.960)	55.414 (40.254)	60.502 (53.613)	51.447 (42.579)
2nd segment (N=24)	1045.1 (3232.4)	408.3 (2600.9)	1805.1 (2350.0)	1120.6 (1493.7)	481.3 (1521.2)	441.3 (1687.2)	0.759 (1.814)	17.213 (27.227)	1.709 (32.648)	10.826 (31.896)
3rd segment (N=13)	-8014.1 (9125.7)	-2887.7 (7342.2)	-4484.7 (6849.5)	-4162.1 (3329.1)	-4548.3 (4084.8)	-3429.2 (4576.6)	-2.799 (4.542)	-57.507 (65.005)	-61.100 (71.815)	-48.734 (109.826)
Avg. ability	21.255 (65.578)	8.269 (49.193)	-3.939 (42.950)	17.804 (26.909)	10.606 (29.171)	3.567 (28.291)	0.001 (0.033)	0.227 (0.477)	0.130 (0.550)	0.085 (0.511)
<i>R</i> <sup>2</sup>	<i>0.08</i>	<i>0.04</i>	<i>0.06</i>	<i>0.14</i>	<i>0.07</i>	<i>0.10</i>	<i>0.14</i>	<i>0.08</i>	<i>0.08</i>	<i>0.07</i>

*Note:* Based on information from 49 teams. In panel A average and SD of ability reflect at the team level respectively average score and standard deviation in scores on Raven's advanced progressive matrices test. In panel B average and ratio of ability reflect at the team level respectively average score and ratio of the maximum to the minimum score on Raven's advanced progressive matrices test. In panel C average and CV of ability reflect at the team level respectively average score and coefficient of variation in scores on Raven's advanced progressive matrices test (with cutoffs at a coefficient of variation in test scores of 0.15 and 0.30). OLS, Median and Robust refer to the estimation method. Median and robust specifications for positive profits are excluded since this variable is dichotomous. Bootstrapped standard errors in parentheses (1000 replications). \*\*\*/\*\*/\* denotes significance at the 1%/5%/10%-level.

**Appendix Table A3.** Ability dispersion, dismissals and team performance

<b>A: Dismissals</b>	Number			Incidence						
	OLS	Median	Robust	OLS						
CV ability	-14.140*** (4.154)	-18.227** (8.399)	-13.057** (5.523)	-8.416*** (3.260)						
CV ability <sup>2</sup>	29.525*** (8.136)	38.475** (17.091)	28.681*** (10.264)	17.191*** (6.489)						
Avg. ability	0.018 (0.027)	-0.001 (0.020)	-0.007 (0.028)	0.012 (0.024)						
<i>Minimum</i>	<i>0.24</i>	<i>0.24</i>	<i>0.23</i>	<i>0.24</i>						
<i>R<sup>2</sup></i>	<i>0.37</i>	<i>0.03</i>	<i>0.36</i>	<i>0.25</i>						
<b>B: Performance</b>	Sales			Profits			Pos. profits	Profits per share		
	OLS	Median	Robust	OLS	Median	Robust	OLS	OLS	Median	Robust
Dismissals (number)	-572.6*** (189.1)	-283.5 (202.5)	-278.9* (164.4)	-193.0** (90.5)	-176.0 (127.4)	-185.2* (98.8)	-0.209 (0.129)	-3.270* (1.775)	-4.392** (1.988)	-3.120* (1.874)
<i>R<sup>2</sup></i>	<i>0.13</i>	<i>0.05</i>	<i>0.10</i>	<i>0.07</i>	<i>0.05</i>	<i>0.10</i>	<i>0.07</i>	<i>0.07</i>	<i>0.06</i>	<i>0.08</i>
Dismissals (incidence)	-774.2*** (269.1)	-355.2 (285.5)	-430.0** (208.7)	-244.3** (123.7)	-164.0 (166.8)	-223.4* (127.8)	-0.259 (0.166)	-4.653** (2.033)	-4.057 (2.768)	-4.038* (2.137)
<i>R<sup>2</sup></i>	<i>0.14</i>	<i>0.06</i>	<i>0.14</i>	<i>0.06</i>	<i>0.03</i>	<i>0.09</i>	<i>0.06</i>	<i>0.08</i>	<i>0.04</i>	<i>0.08</i>

*Note:* Based on information from 49 teams. Average and CV of ability reflect at the team level respectively average score and coefficient of variation in scores on Raven's advanced progressive matrices test. All specifications control for team size. OLS, Median and Robust refer to the estimation method. Median and robust specifications for incidence of dismissals and positive profits are excluded since these variables are dichotomous. Bootstrapped standard errors in parentheses (1000 replications). \*\*\*/\*\*/\* denotes significance at the 1%/5%/10%-level.