methods, data, analyses | Vol. 8(2), 2014, pp. 223-242 DOI: 10.12758/mda.2014.009

The Use of Respondent Incentives in PIAAC: The Field Test Experiment in Germany

Silke Martin[1], *Susanne Helmschrott*[2] & *Beatrice Rammstedt*[1]

[1] *GESIS – Leibniz Institute for the Social Sciences*
[2] *University of Mannheim*

Abstract

In PIAAC, each participating country was required to attain a response rate of at least 50%, as long as evidence was provided that there was either no or only low nonresponse bias in the data. Achieving 50% is a challenge for face-to-face surveys in most Western countries and also in Germany. Previous research showed that the use of incentives is an effective tool to increase response rates in different kinds of surveys. However, incentives may have differential effects on certain socio-demographic groups, because the perceived benefits of an incentive are subjective. To assess the effects of incentives on response rate and nonresponse bias, an experiment with three incentive treatments (€10-coin, €25 and €50 in cash) was implemented in the German PIAAC field test. Results show that response rates increased as the incentive increased. With regard to nonresponse bias, the results are less explicit. According to logistic regressions, the main factors for participation in the €50 condition are age, citizenship, and municipality size and in the €25 condition, only municipality size. Bivariate analyses put these results into perspective. For all treatment groups, a low potential for bias is visible, and there is no statistical evidence that response distributions of the realized sample across treatments are different.

Keywords: Incentive, response rate, experiment, sample composition, PIAAC

© The Author(s) 2014. This is an Open Access article distributed under the terms of the Creative Commons Attribution 3.0 License. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

1 Introduction

The Programme for the International Assessment of Adult Competencies (PIAAC) aimed at producing a high-quality database with reliable and comparable data across the participating countries. Achieving a high response rate was one central indicator for quality in PIAAC. As defined in the international PIAAC technical standards and guidelines, and in accordance with similar cross-national studies, such as the European Social Survey (ESS; Koch, Fitzgerald, Stoop, Widdop, & Halbherr, 2012), the target response rate for each country was set to 70% (OECD, 2010a). Response rates of between 50% and 70% were typically accepted, as long as evidence was provided that there was either no or only low nonresponse and undercoverage bias in the data. Countries not meeting the minimal response rate requirement of 50% were usually not included in the international data set and reports (OECD, 2010a).[1]

Passing the benchmark of a minimum response rate of 50% was a challenge for several countries in PIAAC, because non-participation in large-scale face-to-face surveys is a growing concern all over the world (Atrostic, Bates, Burt, & Silberstein, 2001; Couper & de Leeuw, 2003; de Leeuw & de Heer, 2002; Dixon & Tucker, 2010). As Blohm and Koch (2013) report, for example, four of the 27 countries in the 2010 round of the ESS (European Social Survey, 2012) and eight of the 27 countries participating in 2011 in the European Quality of Life Survey (Eurofound, 2012) failed to reach response rates of 50%.

A serious and constant decrease in response rates for registry-based surveys is also clearly detectable in Germany. National probability surveys in Germany, such as the German General Social Survey (Bevölkerungsumfrage der Sozialwissenschaften, ALLBUS), have experienced a decline in response rates throughout recent years: from approximately 54% in 1994 (Koch, Gabler, & Braun, 1994) to 34% in 2010 (Wasmer, Scholz, Blohm, Walter, & Jutz, 2012). Analogously, in the first round of the ESS, Germany achieved a response rate of approximately 56% (European Social Survey, 2002), whereas in the last two rounds 5 and 6, response

1 Countries were only to be included if analyses indicated that the potential bias is not greater than a potential bias introduced by a response rate between 50% and 70%.

Direct correspondence to
Silke Martin, GESIS – Leibniz Institute for the Social Sciences,
P.O. Box 12 21 55, 68072 Mannheim, Germany
E-mail: silke.martin@gesis.org

Acknowledgment: The authors sincerely thank Michael Blohm and Achim Koch for their valuable input in the decision process identifying an appropriate incentive strategy in PIAAC.

rates of only 31% and 34%, respectively, were realized (European Social Survey, 2012, 2013).

In addition to achieving a substantial response rate, keeping the nonresponse and undercoverage bias negligible, or at least low, was a second major quality criterion in PIAAC (OECD, 2010a). If non-participation in a survey follows a systematic pattern, such that certain groups of sampled persons are less likely to participate than others, nonresponse may cause bias in the data and thus have an impact on the quality of the data (Groves, 2006). Offering an incentive could have a differential effect on the propensity to participate for certain groups and can thus either introduce or reduce nonresponse bias.

Given these standards for PIAAC, the recommendation to use incentives was explicitly embedded in the PIAAC technical standards and guidelines and countries were encouraged to adopt an incentive strategy for improving response rates (OECD, 2010a). The vast majority of the participating countries used some form of incentive or a selection of several incentives during the PIAAC field test (OECD, 2010b). However, only in five countries, Denmark, Germany, Norway, the United Kingdom, and the United States, an incentive experiment was implemented. The incentive experiment in Germany employed three different monetary incentive conditions and aimed to identify the most suitable incentive strategy for the main study.[2]

2 Past Research on Incentives

Previous studies have demonstrated that response rates increase when incentives are provided (e.g., Börsch-Supan, Krieger, & Schröder, 2013; Church, 1993; Singer, 2002; Singer & Kulka, 2002; Singer, Van Hoewyk, Gebler, Raghunathan, & McGonagle, 1999; Singer & Ye, 2013). In his meta-analysis, Church (1993) analyzed 38 mail surveys that have commonly used monetary and non-monetary incentives over the last decades, and concluded that, overall, incentives have a positive effect on the response rate. In particular, the results showed that (a) prepaid incentives work better than conditional incentives, (b) monetary incentives are more effective than non-monetary incentives, and (c) response rates increase with the monetary value of the incentive. In face-to-face or telephone surveys, the effectiveness of incentives has been investigated less (Blohm & Koch, 2013). The most prominent study in this context is the meta-analysis by Singer et al. (1999) of 39 incentive experiments conducted in interviewer-administered surveys in the United States and Canada. They verified that the previously identified effects of incentives

2 PIAAC in Germany was funded by the Federal Ministry of Education and Research (BMBF) with the participation of the Federal Ministry of Labor and Social Affairs (BMAS).

on response rates, although generally smaller than in mail surveys, are also present in face-to-face and telephone surveys. Further international research on incentive experiments conducted in face-to-face or telephone surveys, some of them panel surveys, more or less replicated these findings (e.g., Castiglioni, Pforr, & Krieger, 2008; Eyerman, Bowman, Butler, & Wright, 2005; Jäckle & Lynn, 2008; Rodgers, 2011; Schröder, Saßenroth, Körtner, Kroh, & Schupp, 2013; Singer & Kulka, 2002).

For cross-sectional large-scale assessment surveys, like PIAAC, there is only limited published evidence on the use of incentives or on incentive experiments, to date. This type of survey places some large burdens on respondents because, in addition to a long interviewer-administered interview, respondents have to complete a cognitive assessment on their own. Incentives can be a helpful tool to compensate for this additional burden. In the two central international adult assessment surveys that precede PIAAC, the International Adult Literacy Survey (IALS) and the Adult Literacy and Life Skills Survey (ALL), the use of monetary incentives was prohibited (Murray, Kirsch, & Jenkins, 1997; Statistics Canada, 2002). In IALS, however, Sweden and Germany deviated slightly from this guideline and offered small symbolic incentives. In ALL 2003, the United States provided a conditional incentive of $35 (Krenzke, Mohadjer, & Hao, 2012).

Berlin et al. (1992) and Mohadjer et al. (1997) reported results from an incentive experiment implemented in the 1992 National Adult Literacy Survey (NALS; U.S. Department of Education, 2001) that are in line with the literature and show that incentives significantly increase the response rates. Results from the other PIAAC field test experiments go into the same direction. In the United Kingdom, vouchers (worth £20 or £30) were offered and results showed a significant difference in response rate in favor of the higher incentive (Department for Business Innovation & Skills, 2013).[3] In the United States two incentive conditions ($35 and $50) were concurrently tested. Krenzke et al. (2012) showed that the refusal rate was significantly lower in the $50 condition.

Incentives may have differential effects on certain socio-demographic groups because the perceived benefits of an incentive are subjective and therefore could affect the sample composition (Singer & Kulka, 2002). The effect may be positive and reduce nonresponse bias, when incentives draw individuals into the sample who would otherwise be more prone to refuse (Singer & Ye, 2013). Only a few studies have investigated the effects of incentives on sample composition and response distributions, to date (e.g., Blohm & Koch, 2013; Eyerman et al., 2005; McGrath, 2006; Nicolaas & Stratford, 2005; Singer, 2002; Singer et al., 1999). In summary, these studies provide mixed results. Whereas some studies found no (major) effects of providing incentives on the sample composition (e.g., Blohm & Koch, 2013; Eyerman et al., 2005), other studies report evidence that offering an incentive sup-

3 Results of the Danish and the Norwegian experiment are not available, to date.

ported the recruitment of respondent groups into the sample that otherwise would be underrepresented in the survey, such as e.g., low-income or minority respondents (e.g., Singer, 2002; Singer et al., 1999). In the 1992 NALS, Berlin et al. (1992) found evidence for self-selection of better-educated and wealthy people into the zero-incentive condition, resulting in an overestimation of the population's literacy level in this treatment group.

In the German context, Pforr et al. (forthcoming) have compiled information on ten incentive experiments conducted in eight large-scale face-to-face surveys (two cross-sectional surveys, ALLBUS and PIAAC, and six panel surveys[4]). Given the variation in study and experimental design of these eight surveys, the findings always only refer to some of the analyzed surveys. Pforr et al. found evidence that incentives increase response and retention rates and demonstrated that an increase of the monetary incentive value results in a higher response rate. Cash incentives were more effective than lottery tickets. Mixed results were found with regard to the effects of incentives on nonresponse bias. For several socio-demographic variables, the variable distributions across experimental conditions were analyzed. The results for cross-sectional face-to-face studies indicated that incentives did not affect the sample composition for the selected variables, whereas, for some panel studies, evidence emerged that some groups of respondents were more attracted by incentives than others.

The present paper aims to investigate two central questions in the context of the German PIAAC field test experiment: Do incentives have a positive effect on the response rate? Is there a differential effect of incentives on the sample composition and response distribution?

3 Method

The PIAAC field test had the function of a dress rehearsal for the main study and aimed to define and evaluate, amongst other key aspects, sampling, interviewer training, and survey operation procedures. In Germany, all procedures were implemented as closely as possible to the PIAAC main study parameters. However, given a shorter data collection period than in the main study,[5] some of the main study fieldwork measures (e.g., refusal conversion in re-issue phases or tracing addresses of sampled persons that had moved) could not be realized in the field test.

A registry-based, three-stage stratified and clustered sample design was implemented for the PIAAC field test in Germany, and a gross sample of 3,455 persons

4 German Internet Panel, National Educational Panel Study, German Family Panel, Panel Study "Labor Market and Social Security", Survey of Health, Aging and Retirement, and the Socio-Economic Panel.

5 In Germany, field test data collection took place from April to June 2010.

was selected. In order to depict a representative distribution of small, medium, and large municipalities in Germany, but on a smaller scale, the field test was conducted in only five federal states: Bavaria, Hamburg, Saxony, Schleswig-Holstein, and Thuringia (for more information see Zabal et al., 2014). To obtain a sufficient number of selected persons per federal state, the sampling occurred disproportionately, with oversampling Hamburg and Schleswig-Holstein and selecting fewer cases in Bavaria.

The PIAAC field test in Germany employed three monetary incentives: a €10 commemorative silver coin, engraved with a motif of the 2006 Soccer World Cup, €25 in cash, and €50 in cash. They were randomly allocated to individuals in the gross sample within each sample point with the ratio of 20:40:40.[6] In general, a sample point was allocated to one single interviewer, which ensured that each interviewer worked in all three incentive conditions.

Given the fact that the incentive experiment was not an independent scientific endeavor, the experimental design had a clear limitation. The PIAAC interview, consisting of an interviewer-administered background questionnaire and a self-administered cognitive assessment part, had an average duration of 1 hour and 40 minutes. The decision to use a €10-coin as the baseline was made to account for this substantial interview burden. Thereby, however, we were not able to analyze effects of an incentive compared to a zero-incentive condition. The analyses of the incentive experiment were based on a gross sample of 3,383 eligible cases and a net sample of 1,183 cases (unweighted counts).[7]

Sampled individuals were informed about the survey and the incentive through an advance letter that was sent to them prior to the first interviewer contact. Similarly, interviewers knew which incentive amount was assigned to each sampled individual and could use this information deliberately as a door-opener. Interviewers were instructed to provide the incentive to respondents at the end of an interview.

In contrast to other response rate calculations standards, such as defined by AAPOR (The American Association for Public Opinion Research, 2011), the response rate in PIAAC is a product of the background questionnaire response rate and the cognitive assessment response rate (cf. Mohadjer, Krenzke, & Van de Kerckhove, 2013, p. 12). In accordance with this definition, response rate analyses were calculated by counting full interviews and refusals of the assessment in the numerator as completed cases[8] and subtracting ineligibles and impairments from

6 Overall, the allocation of incentives with this ratio had been implemented successfully across the treatments.

7 Six cases were excluded, because respondents received €50 instead of the pre-assigned incentive amount.

8 We deviated from the completed case definition (see Mohadjer et al., 2013) by excluding literacy-related nonrespondents.

the group of sample persons in the denominator. The following dispositions were summarized as ineligibles and impairments: Death, sample person moved (a) into institution or (b) outside country, hearing and blindness/visual impairment, physical and other disability.

In the PIAAC field test, a proxy variable of proficiency was calculated for each respondent, instead of producing a set of plausible values[9] for each skill domain, as in the main study. This proxy variable is a standardized logit score based on a transformation of the proportion of correct responses to the assessment items (PIAAC Consortium, 2010).

Analysis Plan

In order to answer our research questions, we first compared differences in response rates and nonresponse rates across the treatment conditions. Subsequently, we analyzed whether the incentive conditions potentially introduced some bias. We used variables from the sampling frame, such as age (in five categories), gender, citizenship (in two categories: German and other), and municipality size in three categories (large, medium, and small) that were available for both respondents and nonrespondents, and ran logistic regressions with response as the dichotomous dependent variable for each incentive condition separately. We decided to not include any data from the interviewer case folders or a consumer marketing database (Microm) in these analyses. Although they are, in general, available for all eligible units, they have quality limitations. Case folder information is subject to measurement error, because in the field test, information was not collected in the standardized way like in the main study. Microm variables do not reflect individual case-wise information, but are rather aggregated (information from up to 500 households is combined) and some have a substantial amount of missing data, most probably because sampled addresses could not be categorized.

In a next step, we looked at response distributions of several socio-demographic variables for each treatment group and compared them to the corresponding data from the German 2008 Microcensus, provided by the Federal Statistical Office. We used 2008 Microcensus data because, in 2010, when we first analyzed the experimental data to make a decision for the main study incentive, these were the most current official and available data at that time.

Additional analyses focused on the effects of incentives on the sample composition by comparing response distributions across the incentive treatments, using Chi-Square-Tests of Independence and propensity weighting.

---

9 For definition and computation of plausible values see Yamamoto, Khorramdel, & Von Davier, 2013.

4 Results

To analyze the extent by which response rates increase when a monetary incentive is provided, response rates of the three incentive treatments were compared by means of Chi-Square-Tests of Independence. Table 1 shows the response and nonresponse rates for the overall sample as well as for each treatment group separately. The nonresponse rate is split into nonresponse due to refusal, non-contact, address-related issues, and other reasons.

In the €50 condition, the achieved response rate was 41.7%, compared to 35.4% in the €25 treatment and 26.5% in the €10-coin group. All differences are significant. Even though the PIAAC target response rate of 50% is not achieved for any of the treatment groups, the results clearly demonstrate an increase of the response rate with increasing incentive size.

In general, nonresponse was particularly due to refusals (41.1%, for the overall sample, compared to 22.7% for the remaining reasons). While the response rate increased from lowest to highest incentive amount, the refusal rate developed in the opposite direction: the higher the incentive, the lower the refusal rate. The refusal rates for both the €25 and the €50 condition differed significantly in comparison to the €10-coin group ($p<.01$ and $p<.001$, respectively). Further, the rates for non-contacts, address-related issues, and other reasons for non-participation were also slightly lower in the €50 condition, but these differences did not reach statistical significance.

The second research question addresses the aspect of selectivity in response across treatment groups and differential effects on the sample composition. At first, effects of socio-demographic frame variables on response behavior (1 = response; 0 = nonresponse) were tested for each treatment group separately by means of logistic regressions with the following explanatory variables from the frame:

(a) Age: 16-25 (reference category)/26-35/36-45/46-55/56-65;

(b) Gender: Male (reference category)/female;

(c) Citizenship: German (reference category)/other;

(d) Municipality size: Large with 100,000 or more inhabitants (reference category)/medium with 20,000 to under 100,000 inhabitants/small with under 20,000 inhabitants.

Distributions of the explanatory variables, separately for respondents and nonrespondents, are given in Table A1.1 in the Appendix. Results of the logistic regressions are summarized in Table 2 and indicate no significant effects for the €10-coin incentive group. For both the €25 and the €50 condition, the results demonstrated that individuals living in small municipalities have a significantly higher propensity to participate, compared to individuals residing in large municipalities ($p<.001$). In the €50 condition, this effect was also found for sampled persons living in medium

Table 1 Response and nonresponse rates by incentive treatment

	Overall (n = 3,381)	€10-coin (n = 660)	€25[a] (n = 1,374)	€50[b/c] (n = 1,347)
	%	%	%	%
Response rate	36.2	26.5	35.4***	41.7***/**
Refusal rate	41.1	48.6	40.6**	37.9***/n.s.
Non-contact rate	8.4	9.9	8.5n.s.	7.4n.s./n.s.
Nonresponse rate - address issues	6.8	8.2	6.8n.s.	6.2n.s./n.s.
Nonresponse rate - other reasons	7.5	6.8	8.7n.s.	6.8n.s./n.s.

Notes: Number of cases = eligible sample. To account for disproportionality in sampling, data are adjusted by a correction factor.
a χ^2-Test for comparison of €10-coin and €25
b χ^2-Test for comparison of €10-coin and €50
c χ^2-Test for comparison of €25 and €50
* = $p < .05$, ** = $p < .01$, *** = $p < .001$, n.s. = not significant

municipalities (p<.01). In addition, the €50 incentive seemed to be more attractive for younger individuals and persons with German citizenship. The 36-45 (p<.01), 46-55 (p<.05), and 56-65 (p<.01) age-groups responded significantly less often than the 16 to 25 year-olds. In the €50 treatment, citizenship also had an effect on participation; individuals with non-German citizenship had a lower propensity of providing an interview, but this result was only significant at the 5%-level. While the pseudo R^2 in the €50-model is the highest across all models, overall, the values of the pseudo R^2 are low for all models, indicating only a weak explanation of response behavior through the independent model variables. In addition, significant correlations (p<.01) showed only low strengths between response status and municipality size in the €25 condition (r=-.116) and between response status and age (r=-.085), citizenship (r=.086) and municipality size (r=-.103) in the €50 condition.

In a second step, we compared the response distributions of central sociodemographic variables, for each incentive condition separately, with the corresponding distributions from the German 2008 Microcensus.[10] The response distributions for several frame and survey-relevant outcome variables, such as highest school qualification and employment status, are given in Table 3. With regard to

10 When comparing response distributions with reference data, differences are not only induced by nonresponse bias, but can be due to other error sources (e.g., noncoverage or sampling). The noncoverage rate was low (cf. Zabal et. al, 2014, for main study), and sampling bias is expected to be low, due to probability sampling.

Table 2 Logistic regression of response behavior on frame variables for each incentive treatment

	€10-coin		€25		€50	
	β	*SE*	*β*	*SE*	*β*	*SE*
Gender						
Male (*ref. cat.*)						
Female	-.292	(.183)	-.071	(.116)	-.036	(.115)
Age						
16 to 25 (*ref. cat.*)						
26 to 35	-.073	(.291)	.007	(.194)	-.082	(.193)
36 to 45	-.281	(.285)	-.223	(.182)	-.479**	(.178)
46 to 55	-.376	(.289)	-.172	(.187)	-.391*	(.178)
56 to 65	-.202	(.310)	-.316	(.198)	-.521**	(.190)
Citizenship						
German (*ref. cat.*)						
Other	-.340	(.378)	-.159	(.223)	-.596*	(.232)
Municipality size (No. of inhabitants)						
100,000+ (*ref. cat.*)						
20,000 to <100,000	-.068	(.294)	.105	(.188)	.535**	(.178)
<20,000	.254	(.208)	.565***	(.135)	.474***	(.131)
Constant	-.702**	(.270)	-.636***	(.174)	-.209	(.173)
n	618		1,289		1,282	
Pseudo R^2	0.020		0.027		0.040	

Notes: To account for disproportionality in sampling, data are adjusted by a correction factor.
* = $p < .05$, ** = $p < .01$, *** = $p < .001$

gender, the samples of all treatments included more men than women, when compared to the Microcensus. Whereas the distribution is fairly close to the reference data for both of the cash alternatives, the difference in the €10-coin distribution is obvious. This could be due to the motif of the €10-coin, which was related to the 2006 Soccer World Cup and might have been more attractive to male individuals.

Similar to the effects observed in the multivariate analyses, it can be seen that the €50 incentive was more attractive for the youngest age group. However, at the bivariate level, the proportion of 16 to 25 year-olds is only slightly higher, compared to the Microcensus data, and the share of 36 to 45 year-olds is slightly smaller. The distribution of age in the €25 condition shows the best fit with the Microcensus data, while there are some minor deviations from the expected distri-

Table 3 Comparison of survey estimates with German Microcensus data 2008

	€10-coin %	€25 %	€50 %	MC 08 %
Gender				
Male	57.7	51.9	52.4	50.7
Female	42.3	48.1	47.6	49.3
Age				
16 to 25	18.9	18.1	21.4	18.2
26 to 35	21.7	18.9	18.3	18.3
36 to 45	22.9	24.5	21.7	24.2
46 to 55	20.6	22.0	22.4	21.9
56 to 65	16.0	16.5	16.2	17.5
Citizenship				
German	94.3	93.2	94.7	91.4
Other	5.7	6.8	5.3	8.6
Municipality size (No. of inhabitants)				
100,000+	30.3	26.7	26.4	29.1
20,000 to <100,000	12.6	13.0	17.3	17.3
<20,000	57.1	60.3	56.3	53.6
Highest school qualification				
Low	32.4	25.7	28.4	33.7
Medium	39.9	37.1	39.5	37.7
High	27.7	37.1	32.1	28.6
Employment status				
Employed	80.3	74.5	75.2	70.2
Unemployed	4.0	3.9	5.2	8.3
Inactive	15.6	21.6	19.6	21.5

Notes: To account for disproportionality in sampling, data are adjusted by a correction factor. Microcensus estimates are based on data for the target group of 16 to 65 year olds in Bavaria, Hamburg, Saxony, Schleswig-Holstein, and Thuringia.

bution in the €10-coin condition. Altogether, there is no indication that any of the three distributions of age clearly deviates from the Microcensus distribution.

At the bivariate level, it can be seen that each of the three incentives attracted more target persons with German than with non-German citizenship into the sample, although the effect of citizenship on response behavior in the logistic regression model only reached statistical significance in the €50 condition. Overall, the €25 condition had a slightly better distribution than the €50 condition or the €10-coin group, in comparison to the reference data.

Results observed in the multivariate analysis for municipality size were also visible in the bivariate analysis. Distributions across categories of the variable municipality size showed some deviations from Microcensus distribution in all incentive treatments. Whilst in the €10-coin group, the proportion of persons living in large municipalities was closest to official data, in the €50 condition, the proportion of persons living in medium municipalities matched the Microcensus data perfectly. Altogether, the observed distribution in the €25 treatment deviated clearly from the Microcensus distribution, mainly because the share of residents in medium municipalities is considerably lower and the share of residents in small municipalities considerably higher than in the Microcensus.

Regarding educational attainment, measured as the highest German general school leaving qualification obtained, the €10-coin group revealed a distribution that closely followed the Microcensus distribution, whilst both the €25 and the €50 conditions differed, in comparison to the Microcensus. However, a comparison of these two conditions reveals that the response distribution in the €50 condition was closer to the reference data than the response distribution in the €25 condition, mainly due to a considerable underrepresentation of persons with a low educational level and an overrepresentation of persons with a high educational level in the €25 group.

Next to educational attainment, employment status is considered as a central outcome variable in PIAAC, because skills and employment status are closely linked (Klaukien et al., 2013; OECD, 2013). The distribution of employment status differed considerably from the Microcensus distribution in each treatment group. Particularly in the €10-coin treatment, employed individuals are overrepresented, whereas unemployed and inactive persons are underrepresented.

In order to investigate differential effects of incentives on the sample composition, we analyzed differences in the response distributions for a range of variables across treatment groups by using Chi-Square-Tests of Independence. Results summarized in Table 4 indicate that neither the €25, nor the €50 condition revealed any significant differences in the response distributions for any of the variables, when compared to the €10-coin treatment or to one another.

In addition, we investigated if the incentive treatments differed in the mean outcome variable, the proxy of proficiency. This logit score in the German PIAAC net sample ranges from -4.5360, a value that indicates a lower proficiency level, to 2.7591, a value that represents a higher skill level. Given an average of -.1475 (with a standard deviation of 1.1110) for the overall sample, all of the corresponding logit score means in the three treatment groups were fairly close to this average. While the logit score means of the €10-coin and the €25 treatment (-.1216 and -.1279, respectively; see Table 4) were slightly higher, the logit score mean in the €50 condition was slightly lower (-.1726). Results of the *t*-test for independent samples, however, revealed no significant differences between the treatment groups.

Table 4 Comparison of survey estimates across incentive treatments

	€10-coin %	€25 %	€50 %
Gender (χ^2-*Test*)		(n.s.)	(n.s./n.s)
Male	57.7	51.9	52.4
Female	42.3	48.1	47.6
Age (χ^2-*Test*)		(n.s.)	(n.s./n.s)
16 to 25	18.9	18.1	21.4
26 to 35	21.7	18.9	18.3
36 to 45	22.9	24.5	21.7
46 to 55	20.6	22.0	22.4
56 to 65	16.0	16.5	16.2
Citizenship (χ^2-*Test*)		(n.s.)	(n.s./n.s)
German	94.3	93.2	94.7
Other	5.7	6.8	5.3
Municipality size (χ^2-*Test*)		(n.s.)	(n.s./n.s)
100,000+	30.3	26.7	26.4
20,000 to <100,000	12.6	13.0	17.3
<20,000	57.1	60.3	56.3
Highest school qualification (χ^2-*Test*)		(n.s.)	(n.s./n.s)
Low	32.4	25.7	28.4
Medium	39.9	37.1	39.5
High	27.7	37.1	32.1
Employment status (χ^2-*Test*)		(n.s.)	(n.s./n.s)
Employed	80.3	74.5	75.2
Unemployed	4.0	3.9	5.2
Inactive	15.6	21.6	19.6
	Mean	Mean (n.s.)	Mean (n.s./n.s)
Proficiency proxy (*t-Test*)	-.1216	-.1279	-.1726

Notes: To account for disproportionality in sampling, data are adjusted by a correction factor. n.s. = not significant

Table 5 Comparison of survey estimates for non-propensity and propensity weighted data across incentive treatments

	€10-coin		€25		€50	
	non-propensity weighted	propensity weighted	non-propensity weighted	propensity weighted	non-propensity weighted	propensity weighted
	%	%	%	%	%	%
Highest school qualification						
Low	32.4	31.8	25.7	25.0	28.4	27.3
Medium	39.9	40.2	37.1	36.2	39.5	39.3
High	27.7	28.0	37.1	38.8	32.1	33.4
Employment status						
Employed	80.3	80.8	74.5	73.7	75.2	75.4
Unemployed	4.0	3.8	3.9	4.0	5.2	5.3
Inactive	15.6	15.3	21.6	22.3	19.6	19.3
	Mean	Mean	Mean	Mean	Mean	Mean
Proficiency proxy	-.1216	-.1451	-.1279	-.1201	-.1726	-.1922

In order to simulate a sample distribution that would have resulted if all sample persons had participated, we finally weighted the data with propensity weights that accounted for differential response behavior and which were computed in the logistic regression models for each incentive treatment separately. Distributions of propensity weighted variables were further compared to the distributions of variables without propensity weights (see Table 5). Considerable differences in the distributions would be an indicator that differential response behavior has an effect on the sample composition.

As Table 5 depicts, propensity weights had hardly any effect on the distribution of educational attainment in any of the incentive treatment groups. With regard to employment status, there was no effect on the distribution from propensity weighting for the €10-coin and the €50 condition. In the €25 treatment, however, the share of employed persons was slightly reduced and the share of inactive individuals increased slightly through the weighting.

Propensity weighting in both the €10-coin and the €50 condition resulted in somewhat lower means of the proxy variables for proficiency, whereas the average proficiency score in the €25 condition became slightly higher. Given a range of 7.2951 for this variable, these changes, however, can be considered negligible.

5 Discussion

Previous studies on the use of incentives showed that incentives have a positive effect on response rates. Only few studies, however, investigated effects of incentives on the sample composition and response distributions. In particular, there is only limited published evidence on the use of incentives in adult assessment surveys. In the present study, we analyzed results from the experiment of testing three incentive conditions in the German PIAAC field test. Two central questions were addressed: Do incentives have a positive effect on the response rate? Is there a differential effect of incentives on the sample composition and response distribution?

Results from the response rate analysis of this experiment are straightforward. As expected from the literature, we found that incentives are an effective tool for increasing the response rate. For the PIAAC field test incentive experiment, we observed that response rates increased significantly with increasing amounts of the incentive.

With regard to the second research question the results are less explicit. While results of the multivariate analyses indicate a potential for bias in the €25 and €50 condition for municipality size, these results are put into perspective, at least partly for age and citizenship, based on the bivariate analyses, e.g., by comparing response distributions to Microcensus data or across incentive treatments. Response distribution of citizenship and municipality size differ across all treatment groups when compared to the reference data. Results thus indicate that non-German individuals and persons who live in large municipalities have, in general, a lower response propensity.

The €50 incentive was, however, more attractive for 16 to 25 year-olds. This effect is significant in the logistic regression and results in a slightly higher proportion of 16 to 25 year-olds, compared to the Microcensus. However, the difference in the response distribution of age in the €50 condition does not reach statistical significance when compared to the response distribution of age in the €25 or €10-coin condition.

For educational attainment, the results reveal that the variable's distribution in the €10-coin group showed the best match with Microcensus data. Both cash alternatives introduced some bias in the data, but compared to the €25 treatment, educational levels are better represented in the €50 condition. With regard to employment status, none of the response distributions of any incentive treatment matched the distribution of the Microcensus data well.

In general, comparisons of response distributions of central socio-demographic variables and of the mean logit scores across all incentive conditions did not provide evidence that the incentive size changed the sample composition in any treatment group in a substantial way. In contrast to results reported by Berlin et al. (1992) for the NALS survey in 1992, findings from the German PIAAC field test do

not confirm the observation that there is a self-selection of more educated or more skilled individuals in the condition with the smallest monetary amount (€10-coin) and, thus, an overestimation of the proficiency level in this treatment group. However, the results are not perfectly comparable, because in the NALS experiment, the control group received no incentive at all. Moreover, in the PIAAC field test data, only an approximation of proficiency was used as indicator.

Finally, by using propensity weights, obtained from the logistic regression, we see that the differential effects for response hardly changed the response distributions of educational attainment, employment status, and mean proficiency score.

In conclusion, the €50 incentive had the strongest positive effect on the achieved response rate. In this condition, some groups of people had a higher propensity to participate. This had, however, only a minor impact on the sample composition. Moreover, there is a low potential for bias in the data for each treatment group, because response distributions of some variables show minor deviations in each of the treatments, compared to Microcensus data. When response distributions of each treatment were compared with one another, statistical evidence that they are different could not be found.

For future cycles of PIAAC, it would be interesting to assess whether the current findings can be replicated in other participating countries and to which extent different survey operation designs, protocols and procedures (e.g., sampling designs, different types of data collection agencies or fieldwork instructions for interviewers) moderate the results. In the context of large-scale adult assessment surveys, further research on the impact of incentives on final proficiency scores (as computed in terms of plausible values) would be beneficial in order to evaluate potential motivational effects of the incentive amount on the respondent's effort to accomplish the cognitive assessment part.

References

Atrostic, B. K., Bates, N., Burt, G., & Silberstein, A. (2001). Nonresponse in U.S. Government household surveys: Consistent measures, recent trends, and new insights. *Journal of Official Statistics, 17*(2), 209-226.

Berlin, M., Mohadjer, L., Waksberg, J., Kolstad, A., Kirsch, I. S., Rock, D., & Yamamoto, K. (1992). An experiment in monetary incentives. In: American Statistical Association, *Proceedings of the Survey Research Methods Section 1992* (pp. 393-398). Retrieved September 2014 from http://www.amstat.org/sections/srms/Proceedings/allyearsf.html

Blohm, M., & Koch, A. (2013). Respondent incentives in a national face-to-face survey: Effects on outcome rates, sample composition and fieldwork efforts. *methods, data, analyses,* 7(1), 89-122.

Börsch-Supan, A., Krieger, U., & Schröder, M. (2013). *Respondent incentives, interviewer training and survey participation*. SHARE Working Paper Series 12-2013. Munich: Munich Center for the Economics of Aging (MEA).

Castiglioni, L., Pforr, K., & Krieger, U. (2008). The effect of incentives on response rates and panel attrition: Results of a controlled experiment. *Survey Research Methods,* 2(3), 151-158.

Church, A. H. (1993). Estimating the effect of incentives on mail survey response rates: A meta-analysis. *Public Opinion Quarterly,* 57(1), 62-79.

Couper, M. P., & de Leeuw, E. D. (2003). Nonresponse in cross-cultural and cross-national surveys. In J. A. Harkness, F. J. R. van de Vijver & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 157-177). New York: John Wiley & Sons.

de Leeuw, E. D., & de Heer, W. (2002). Trends in household survey nonresponse. A longitudinal and international comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 41-54). New York: John Wiley & Sons.

Department for Business Innovation & Skills. (2013). *The International Survey of Adult Skills 2012: Adult literacy, numeracy and problem solving skills in England - Appendices.* BIS Research Paper No. 139A. London: Department for Business, Innovation and Skills. Retrieved September 2014 from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/248689/bis-13-1221an1-international-survey-of-adult-skills-2012-appendices.pdf

Dixon, J., & Tucker, C. (2010). Survey nonresponse. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of survey research* (2nd ed., pp. 593-630). Bingley: Emerald.

Eurofound. (2012). *Third European Quality of Life Survey – Quality of life in Europe: Impacts of the crisis.* Luxembourg: Publications Office of the European Union.

European Social Survey. (2002). *ESS1-2002 Documentation Report. The ESS Data Archive.* (6.3 ed.). Bergen: Norwegian Social Science Data Services. Retrieved September 2014 from http://www.europeansocialsurvey.org/docs/round1/survey/ESS1_data_documentation_report_e06_3.pdf

European Social Survey. (2012). *ESS5-2010 Documentation Report. The ESS Data Archive.* (3.0 ed.). Bergen: Norwegian Social Science Data Services. Retrieved March 2014 from http://www.europeansocialsurvey.org/docs/round5/survey/ESS5_data_documentation_report_e03_0.pdf

European Social Survey. (2013). *ESS6-2012 Documentation Report. The ESS Data Archive* (1.3 ed.). Bergen: Norwegian Social Science Data Services. Retrieved March 2014 from http://www.europeansocialsurvey.org/docs/round6/survey/ESS6_data_documentation_report_e01_3.pdf

Eyerman, J., Bowman, K., Butler, D., & Wright, D. (2005). The differential impact of incentives on refusals: Results from the 2001 national household survey on drug abuse incentive experiment. *Journal of Economic and Social Measurement,* 30, 157-169.

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly,* 70(5), 646-675.

Jäckle, A., & Lynn, P. (2008). Respondent incentives in a multi-mode panel survey. Cumulative effects on nonresponse and bias. *Survey Methodology,* 34(1), 105-117.

Klaukien, A., Ackermann, D., Helmschrott, S., Rammstedt, B., Solga, H., & Wößmann, L. (2013). Grundlegende Kompetenzen auf dem Arbeitsmarkt. In B. Rammstedt (Ed.), *Grundlegende Kompetenzen Erwachsener im internationalen Vergleich. Ergebnisse von PIAAC 2012* (pp. 127-165). Münster: Waxmann.

Koch, A., Fitzgerald, R., Stoop, I., Widdop, S., & Halbherr, V. (2012). *Field Procedures in the European Social Survey Round 6: Enhancing Response Rates.* Mannheim: European Social Survey, GESIS. Retrieved 16 November 2014 from http://www.europeansocialsurvey.org

Koch, A., Gabler, S., & Braun, M. (1994). *Konzeption und Durchführung der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (ALLBUS) 1994*. ZUMA-Arbeitsbericht 94/11. Mannheim: ZUMA.

Krenzke, T., Mohadjer, L., & Hao, H. (2012). *Programme for the International Assessment of Adult Competencies: U.S. Incentive Experiment*. Paper presented at the 67th Annual Meeting of the American Association of Public Opinion Research, May 2012, Orlando, FL.

McGrath, D. E. (2006). *An incentives experiment in the U.S. consumer expenditure quarterly survey.* Paper presented at the Joint Statistical Meetings, ASA Section on Survey Research Methods, August 2006, Seattle, WA.

Mohadjer, L., Berlin, M., Rieger, S., Waksberg, J., Rock, D., Yamamoto, K., . . . Kolstad, A. (1997). The role of incentives in literacy survey research. In A. C. Tuijnman, I. S. Kirsch & D. A. Wagner (Eds.), *Adult basic skills: Innovations in measurement and policy analysis* (pp. 209-244). Cresskill, NY: Hampton Press.

Mohadjer, L., Krenzke, T., & Van de Kerckhove, W. (2013). Indicators of the quality of the sample data. In OECD (Ed.), *Technical report of the Survey of Adult Skills (PIAAC)* (Chapter 16, pp. 1-30). Paris: OECD. Retrieved March 2014 from http://www.oecd.org/site/piaac/surveyofadultskills.htm

Murray, T. S., Kirsch, I. S., & Jenkins, L. B. (1997). *Adult literacy in OECD countries: Technical report on the first International Adult Literacy Survey.* Washington D.C.: National Center for Education Statistics.

Nicolaas, G., & Stratford, N. (2005). A plea for the tailored use of respondent incentives. In: C. van Dijkum, J. Blasius, & C. Durand (Eds.), Proceedings of Sixth International Conference on Social Science Methodology: Recent developments and applications in social research methodology (pp. 1–15). Amsterdam: Budrich Verlag.

OECD. (2010a). *PIAAC Technical Standards and Guidelines* (December 2010). Retrieved March 2014 from http://www.oecd.org/site/piaac/surveyofadultskills.htm

OECD. (2010b). *Survey Operations Quality Control Field Test Monitoring Report. PIAAC(2010_12)FT_SurveyOperation_QCFinalReport_11Dec10.docx*. Unpublished document.

OECD. (2013). *OECD skills outlook 2013: First results from the Survey of Adult Skills*: Retrieved March 2014 from http://www.oecd.org/site/piaac/Skills%20volume%201%20%28eng%29--full%20v12--eBook%20%2804%2011%202013%29.pdf

PIAAC Consortium. (2010). *Standardized_Logit_File_Description.pdf.* Unpublished document.

Pforr, K., Blohm, M., Blom, A. G., Erdel, B., Felderer, B., Fräßdorf, M., . . . Rammstedt, B. (forthcoming). Are incentive effects on response rates and nonresponse bias in large-scale, face-to-face surveys generalizable to Germany? Evidence from ten experiments. *Public Opinion Quarterly.*

Rodgers, W. L. (2011). Effects of increasing the incentive size in a longitudinal study. *Journal of Official Statistics*, 27(2), 279-299.

Schröder, M., Saßenroth, D., Körtner, J., Kroh, M., & Schupp, J. (2013). Experimental evidence of the effect of monetary incentives on cross-sectional and longitudinal response: Experiences from the Socio-Economic Panel (SOEP). *SOEPpapers on Multidisciplinary Panel Data Research,* No. 603. Berlin: DIW.

Singer, E. (2002). The use of incentives to reduce nonresponse in household surveys. In R. M. Groves, D. A. Dillman, D. A. Eltinge & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 163-177). New York: John Wiley Co.

Singer, E., & Kulka, R. A. (2002). Paying respondent for survey participation. *Survey methodology program, No. 092*.

Singer, E., Van Hoewyk, J., Gebler, N., Raghunathan, T., & McGonagle, K. (1999). The effect of incentives on response rates in interviewer-mediated surveys. *Journal of Official Statistics,* 15(2), 217-230.

Singer, E., & Ye, C. (2013). The use and effects of incentives in surveys. *The ANNALS of the American Academy of Political and Social Science,* 645(1), 112-141.

Statistics Canada. (2002). *Public use microdata file. User's manual*. Ottawa.

The American Association for Public Opinion Research. (2011). *Standard definitions: Final dispositions of case codes and outcome rates for surveys. 7th edition*.

U.S. Department of Education, National Center for Education Statistics (2001). *Technical report and data file user's manual for the 1992 National Adult Literacy Survey.* Washington, DC: National Center for Education Statistics.

Wasmer, M., Scholz, E., Blohm, M., Walter, J., & Jutz, R. (2012). *Konzeption und Durchführung der "Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (ALLBUS) 2010.* GESIS-Technical Reports. Köln: GESIS – Leibniz-Institut für Sozialwissenschaften.

Yamamoto, K., Khorramdel, L., & Von Davier, M. (2013). Scaling PIAAC cognitive data. In OECD (Ed.), *Technical report of the Survey of Adult Skills (PIAAC)* (Chapter 17, pp. 1-33). Paris: OECD. Retrieved March 2014 from http://www.oecd.org/site/piaac/surveyofadultskills.htm

Zabal, A., Martin, S., Massing, N., Ackermann, D., Helmschrott, S., Barkow, I., & Rammstedt, B. (2014). *PIAAC Germany 2012: Technical report*. Münster: Waxmann.

Appendix

Table A1.1 Distributions of socio-demographic variables for respondents and nonrespondents

	€10-coin		€25		€50	
	R	NR	R	NR	R	NR
	%	%	%	%	%	%
Gender	(*n*=175)	(*n*=485)	(*n*=486)	(*n*=888)	(*n*=561)	(*n*=786)
Male	57.7	51.3	51.9	49.4	52.4	50.6
Female	42.3	48.7	48.1	50.6	47.6	49.4
Age	(*n*=175)	(*n*=479)	(*n*=486)	(*n*=876)	(*n*=561)	(*n*=781)
16 to 25	18.9	15.4	18.1	15.7	21.4	15.4
26 to 35	21.7	18.8	18.9	17.8	18.3	15.5
36 to 45	22.9	23.8	24.5	25.3	21.7	25.7
46 to 55	20.6	25.7	22.0	21.7	22.4	23.6
56 to 65	16.0	16.3	16.5	19.5	16.2	19.8
Citizenship	(*n*=175)	(*n*=452)	(*n*=486)	(*n*=823)	(*n*=561)	(*n*=725)
German	94.3	91.6	93.2	90.8	94.7	89.9
Other	5.7	8.4	6.8	9.2	5.3	10.1
Municipality size (No. of inhabitants)	(*n*=175)	(*n*=485)	(*n*=486)	(*n*=888)	(*n*=561)	(*n*=786)
100,000+	30.3	34.2	26.7	36.3	26.4	37.6
20,000 to <100,000	12.6	16.1	13.0	15.9	17.3	13.7
<20,000	57.1	49.7	60.3	47.8	56.3	48.7

Notes: R = respondents; NR = nonrespondents. To account for disproportionality in sampling, data are adjusted by a correction factor.