Barbara Jancewicz

MEASUREMENT OF INCOME INEQUALITY RE-EXAMINED: CONSTRUCTING EXPERIMENTAL TESTS BY QUESTIONNAIRE

1. Introduction

People recognize income inequality as an important topic that influences economies, culture and their lives (Wilkinson and Pickett, 2010). But how to exactly define inequality is a difficult question – in practice we assume definitions imposed by inequality measures (functions that transform data about people's incomes into a single number). Experts propose many different ways of measuring inequality, and there is an ongoing discussion about which of them is best. At the base of the discussions are axioms – propositions stating how a measure should behave under certain circumstances, e.g. when all incomes are raised by 10%. There are many rivalry axioms that portray different ways of defining inequality. Many of the currently used measures, like the Gini, coefficient, Atkinson index and Theil index, share a small set of axioms, some of which are still considered controversial. Furthermore, there is no easy, objective way to decide which axiom is better, since all of them have their merits.

The perceived simplicity of inequality measures makes discussions about inequality easier, both for trained experts and for members of the general public that have started to take an interest in the topic. This rise in popularity has caused a lowered depth of understanding of inequality by the interested parties and, as a consequence, common misinterpretations. Therefore, there is a need to define inequality measures in a way that best corresponds with people's intuition and perception of inequality.

There is a lot of research into the perception of income inequality by ordinary people. For example Decoster and Schokkaert (2002) talk about how the Flemish working population perceives inequality in their country, Cuena et

al. (2004) write about student's perception of inequality in their leaky-bucket experiments. However, very few scientists focus on how the understanding of inequality differs from the meaning of income inequality measures, when looking from the axiomatic perspective. It is the purpose of this article to repeat and critically examine one of the first and broadest studies in this field.

The earliest inquiry in this topic was carried out by Cowell (1985), later in collaboration with Amiel (1992). They conducted an income inequality questionnaire on 1108 students in the USA, the UK, Israel and Germany. While their research has shown that axioms shared by the most popular inequality measures do not have overwhelming support among the respondents, they also found many of their subjects' answers to be illogical, inconsistent or inexplicable. Further research by Amiel and Cowell (1999; 2002) and later by Ballano and Ruiz-Castillo (1993) and Harrison and Seidl (1994) also resulted in low axiom support and multiple puzzling answers. All these studies raised important questions about the validity of the way we measure income inequality. Unfortunately, the field of testing axioms with the population is quite new and scientists are struggling to understand the obtained answers and how to find the best way of eliciting respondents' views on income inequality. Because of this, the methodology used in the studies is still imperfect and the results require further verification.

The aim of this article is to provide insight and suggest improvements to the methodology of research into the perception of inequality. This paper also presents and compares results from a Polish replication of the research originally conducted by Amiel and Cowell. The article concentrates on the first questionnaire created by Amiel and Cowell, since it incorporates a variety of ideas on how to elicit views on such a difficult topic and takes an innovative approach to the problem. Other researchers tried to touch on this issue from this or similar perspectives, with mixed success: Devooght (2003) made similar inquiries, though he based his questionnaire not on axioms, but on the theory of inequality by Temkin (1993); Traub et al. (2003; 2009) asked respondents to rank distributions in different settings, e.g. self-concern and social-planner, and then tried to interpret their answers from the axiomatic perspective. Amiel and Cowell themselves expanded their scope of study by comparing inequality with risk, polarisation, welfare and other concepts (2001; 2002; 2010 with Ramos, 2012 with Gaertner). They avoided the problem of illogical responses by limiting their questionnaire's verbal section and accepted the remaining inconsistencies as normal. They also tried to attribute inconsistencies to respondents' individual characteristics (2004 with Slottje). It is the author's belief that the original methodology, with some

improvements, will yield better results and fewer inconsistent answers from the respondents.

For the purpose of understanding Amiel and Cowells' results and improving the methodology, their study (1992; 1999) was replicated on a group of Polish students. Additionally, many respondents who gave puzzling answers on the initial questionnaire were invited to an in-depth interview to better understand their reasoning. This paper very thoroughly analyses and compares the results obtained and the questionnaire itself.

The paper is structured as follows. Section 2 introduces possible approaches to evaluation of income inequality measures. Section 3 describes the combined quantitative and qualitative approach used in current research. Section 4 compares the results obtained by Amiel and Cowell with those from the Polish inquiry and discusses problems with the questionnaire offering possible solutions. Section 5 concludes.

2. How to evaluate performance of income inequality measures

There are many approaches one can take in order to judge to what extent an inequality measure achieves its goal. The first problem is whose opinion to take into consideration, the second – from what perspective to evaluate these measures.

Who should define what income inequality means? Originally these were economic experts. However, data about income inequality are mostly used in order to guide policies, hence the policymakers' definition should also matter. The policymakers themselves care about inequality mainly due to the fact that citizens care about it and consider it an important factor. Therefore we should also include all the citizens' opinions in our research. Going to these lengths has several major drawbacks that have stopped many scholars from performing such studies so far. Income inequality measures are complicated functions that take the whole distribution of incomes and transform it into a single number. The whole notion of inequality is complicated, thus most approaches to evaluation of income inequality measures require a significant prior knowledge, ability, time and concentration from the evaluator. Consequently, eliciting the views of all citizens, or even a relatively small representative sample, is a demanding task.

The problem of who should define what income inequality means is closely related to the problem of how to define it. The first and most popu-

lar way of evaluating income inequality measures is straightforward: through a measures' equation and the rationale behind it. For example, the Gini coefficient can be described as a mean standardized difference between all incomes, while the Shutz index (sometimes called the Robin Hood index) is the proportion of income that should be redistributed in order to create perfect equality. However, each index can be formulated and interpreted in at least several ways. The aforementioned Gini coefficient has over 20 possible formulations and interpretations related to them. Furthermore, thorough the understanding of an equation and its interpretation with all its consequences requires significant mathematical knowledge and effort. Thus, such an approach to evaluating income inequality measures significantly limits the potential group of evaluators.

Another way to evaluate income inequality measures is through their results. For every pair of income distributions, an inequality measure either considers them both equally unequal, or judges one as more unequal. Thus, we can ask a respondent to evaluate income inequality in a set of examples and check which income inequality measure produces the same ranking. This approach is often used, however it has one significant drawback: the results obtained heavily rely on the choice of specific examples used in the study. Since in one such study we can include several income distributions, only a few properties of income distribution can be tested.

The third option is to base our study on an income inequality measurement framework. There are a few frameworks trying to define income inequality. The most well-known and classic one relies on axioms, rules that an income inequality measure could abide to. These rules are often simple concepts that relate to concrete situations when incomes change. Axioms allow for the dividing of a complicated measure definition into a set of smaller and simpler ones. However, this approach also has its drawbacks. Some sets of reasonable axioms don't lead to any measure at all, because of conflicts between some of them. Other combinations can produce surprising results that the author did not foresee. Most importantly however, we do not know if people's views on income inequality can be expressed through simple axioms. To summarize, approaching the problem of income inequality measures through axioms simplifies research construction, but, just like other approaches, has significant drawbacks.

Nevertheless, the axiomatic approach was the one taken by Amiel and Cowell: each verbal question in their questionnaire describes one situation of changing incomes, and asks a respondent how they think it influenced

inequality; each numerical question presents an application of a certain axiom. Thus, the results and following interpretation is also axiom oriented.

3. Methodology

Problems encountered by Amiel and Cowell in the interpretation of the obtained answers are partially a result of choosing the axiomatic approach and partially due to the fact that they elicited respondents' views in a form of a questionnaire. The problems included: axioms that in certain combinations produced puzzling answers, errors, and self-conflicting responses. Unfortunately, a result of their quantitative interview was simply a list of choices made by respondents, with no possibility of asking additional questions after the study was completed. Although questionnaires often encourage people to leave comments or afterthoughts, few respondents really do. Hence, the task of understanding the reasons and motivations of the people questioned is laid solely on the researchers, with little or no additional guidance from the respondents.

The lack of understanding for the obtained results indicates that we need to take a different approach. In this case, since our main interest is discovering motivations and reasons for certain choices made in a questionnaire, the best method seems to be a qualitative interview. We want to concentrate on people whose responses were described by Amiel and Cowell as 'unconventional', which is why it would be best to interview those who gave such puzzling replies. In order to discriminate such respondents and check whether our sample seems to perceive inequality similarly to Amiel and Cowell's, their study was replicated. The first stage of the research consisted of an auditory questionnaire conducted in the Warsaw School of Economics and in the Institute of Sociology of the University of Warsaw. The questionnaire was an exact translation (done by the author) of the one used in the first Amiel and Cowell research (1992; 1999).

The second part of the research consisted of a series of in-depth interviews with some of the questionnaire respondents. In the first part of the study all the respondents were given a questionnaire with a code assigned to it and a contact card, which they were asked to fill in if they agreed to participate in the second part of the research. Almost one third (31.1%) of the sample agreed to an in-depth interview. From among those who also added their questionnaire's code to the contact card (25.8% of the sample) the interview-

ees were selected[1]. The interviewees (as they will be hereinafter referred to), who were selected, presented almost every type of unconventional or illogical answer, so different problems could be discussed and analysed.

The sample used by Amiel and Cowell consisted of:

> "(…) upper-level undergraduates who had some training in economics but who had not previously taken courses that included studying the measurement of income inequality" (1992, p. 7).

The idea was to use subjects:

> "who are likely to avoid arithmetical mistakes and logical slips" (1992, p. 5).

In an attempt to replicate their research as closely as possible, the Polish sample consisted of undergraduates who did not frequent any inequality measurement classes. In order to ensure arithmetical abilities, all our respondents were students who had some training in statistics as well as the Warsaw School of Economics students who had already taken a basic course in economics[2].

The first study done by Amiel and Cowell showed only small differences between the answers of respondents from different countries who received the questionnaire translated into their native languages. Further research by those authors showed mixed results concerning variations of inequality perceptions among different nationalities, type of education and sex. Nevertheless, the samples used by the researchers were not random, so the results obtained are not truly comparable, hence no statistical tests of significance apply. However, for testing a new area of inquiry and improving methodology, these samples suffice, especially when combined with qualitative methods of research.

4. Results

The results of the Polish study is twofold: on the one hand we have another set of results of the questionnaire that was already conducted in various countries; and on the other hand, thanks to qualitative interviews and

[1] One person filled in the questionnaire with as many as five inconsistencies, two people with four and one person with none (for comparison). Another six interviewees ranged from two to three inconsistencies and were selected so that all possible puzzling answers would be covered.

[2] Students of the Institute of Sociology take a basic economy course slightly later, when they might have already taken an income inequality class.

methodological analysis, we gain a better understanding of the respondents' answers.

This section presents five main topics that were addressed in Amiel and Cowell's study, namely: effect of income transformations, population replication, the principle of transfers, decomposition by population subgroups and unbalanced enrichment of a society. The questionnaire was organized in two sections: the first one consisted of questions using numerical examples of income distributions, while the second one contained verbal questions about rules guiding ones' decisions while an income change occurred. This structure was intended to "(…) allow for the 'learning-by-doing' (…)" (1992, p. 6). This section summarises the results obtained both in the original and the replicated study and also describes the questionnaire itself, but is organized according to a questions' topic instead of the order in which they appeared in the questionnaire.

4.1. Income transformations

Scale and translation invariance are two rivalling axioms that deal with the problem of how to compare distributions with different income means. Scale invariance assumes that inequality remains unchanged when all incomes are multiplied by a positive constant, while translation invariance claims that it is adding a constant that has no effect on inequality level.

In the numerical section respondents were asked to choose a more unequal income distribution out of the two presented to them. In question 1 the second distribution was created from the first one by doubling incomes and in question 2 by adding 5 units to each income. Throughout the article respondents will be labelled supporters of certain axioms if they answer according to those rules of the selected set of questions. For example, a supporter of a scale invariance axiom should mark "the same" in question 1 (that multiplying incomes preserves the level of inequality), and "down" in question 2 (that adding 5 units to each income decreases inequality). Followers of translation invariance should mark "up" in the first question (since multiplying incomes increases absolute differences) and "same" in the second question. Someone can be categorized a supporter of scale invariance when only question 1 is considered, but can be labelled a non-supporter when both questions 1 and 2 are taken into account, for instance when he or she answered "the same" to both questions.

Q1.	A = (5, 8, 10)	B = (10, 16, 20)
Q2.	A = (5, 8, 10)	B = (10, 13, 15)

Table 1 shows that 37% of the respondents in the original study and 16% in the Polish one are supporters of scale invariance, when answers to both numerical questions are considered. In case of translation invariance the corresponding rates are 17% and 26%. Many respondents, especially in the Polish sample (24%), chose to answer "the same" in both questions, which is puzzling since those axioms are considered rivalling. Accepting that both are theoretically possible leads to an enormous amount of distributions that are considered to have the same level of inequality even though they differ greatly. For example distributions like (1000, 1002, 1003) and (1000, 2000, 3000) would be considered equally unequal. Amiel and Cowell commented on their respondents' answers:

> "(…) there is a bias in favour of saying 'the same' even when such a response would appear illogical: perhaps this reflects an innate 'safety first' response on the part of student respondents" (1992, pp. 12–13).

The follow-up interviews clarified this apparent discrepancy to some extent. Many interviewees neither saw the conflict between their answers nor understood the result of accepting two axioms simultaneously. Both pairs of distributions are similar in some way ("So, they are generally the same, but (…), from what I see they differ by 5" [respondent no. 6038]), so when they are considered separately, it encourages the answer "the same" in both of them. Others considered the answer "the same" as equal to "I don't know". From the researcher's point of view, it would be better if respondents that truly had no opinion left the question empty, or analysed the problem more thoroughly. However, interviewees admitted that when in doubt they either marked "the same" or chose a random answer, while leaving the question empty was the option of "last resort". The decision to include an "I don't know" answer is a difficult one and controversial in the methodological literature. Nevertheless, in the theory of income inequality there is an option of non-comparability of some income distributions. Amiel and Cowell decided not to include the "I don't know" option in their study, since they aimed at minimizing non-responses. They were expecting that in the case of an uncomparable pair of examples, the respondents would leave an appropriate comment in their questionnaire.

During the follow-up interviews, we asked respondents who answered "the same" in both questions to rank all three distributions instead of comparing them pairwise: (5, 8, 10), (10, 16, 20), (10, 13, 15). Most interviewees put them at different inequality levels, choosing one axiom over another. What motivated them to do that was the comparison between the distribu-

tions after income changes ((10, 16, 20) and (10, 13, 15)), which was not in the original questionnaire. Thus, inviting respondents to consider both axioms concurrently reduces the amount of puzzling answers.

Table 1

Income transformation – answers to numerical questions compared

A&C		**Add 5 units (q2) (%)**		
		Down	**Up**	**Same**
Double income (q1)	Down (%)	8	**2**	**5**
	Up (%)	15	3	17*
	Same (%)	37**	**5**	**9**

Source: Amiel and Cowell (1992).

Poland (N = 131)		**Add 5 units (q2) (%)**		
		Down	**Up**	**Same**
Double income (q1)	Down (%)	6	**0**	**2**
	Up (%)	15	8	26*
	Same (%)	16**	**2**	**24**

* translation invariance supporters
** scale invariance supporters
Bold: puzzling answers

Source: own research.

Those who answered in accordance with scale or translation invariance in numerical questions did not necessarily agree with those axioms in general. Three verbal questions were asked to elicit views about those axioms – one depicted a situation of income multiplication, the other two of deducting and adding a fixed amount of money to everyone's income.

Q10. Suppose we double the "real income" of each person in a society, when not all the initial incomes are equal.

(a) Each person's share remains unchanged, so inequality remains unchanged.

(b) Those who had more also get more, so inequality has increased.

(c) After doubling incomes more people have enough money for basic needs, so inequality has fallen.

Q11a. Suppose we add the same fixed amount to the incomes of each person in a society, when not all the initial incomes are equal.
(a) Inequality has fallen because the share of those who had more has fallen.
(b) Inequality remains the same.
(c) Inequality has increased.
Q11b. Suppose instead of adding we deduct a fixed amount from each person's income. Then inequality…
(a) Is the same.
(b) Increases.
(c) Decreases.

When asked verbally, 45–47% of the respondents agreed with the scale invariance axiom both in the Polish and in the A&C sample. The level of support for translation invariance is higher among Poles (60% when adding, 50% when deducting income) than among students questioned by A&C (35% and 28% respectively)[3], but differences between verbal questions concerning translation invariance are similar.

Table 2

Income transformations – answers to verbal questions

Sample	N	Double income (q10) (%)			Add fixed sum (q11) (%)			Deduct fixed sum (q11) (%)		
		Down	Up	Same	Down	Up	Same	Down	Up	Same
A&C	1108	12	40	47**	58	6	3%*	7	64	28*
Poland	122–121	15	39	45**	38	2	60*	4	46	50*

* translation invariance supporters
** scale invariance supporters

Source: Amiel and Cowell (1992), own research.

Analysing answers to two verbal questions simultaneously (as presented in Table 3) leads to a significant fall in axioms support: by 17% and 29% for scale invariance in the A&C and Polish sample respectively (to 30% and 16%) and by 18% and 40% for translation invariance (to 17% and 20%).

[3] Polish students answered in favour of translation invariance surprisingly often, possibly studying properties of variance (which is translation invariant) during their basic statistical course could have influenced their judgement.

Again, there are a significant percentage of responses "the same" to all the questions concerning income transformation.

Respondents were asked about support for translation invariance twice: in the case of adding and deducting a fixed amount of money from each income (11a and 11b). Both of these questions referred to the same numerical question (q2), yet the ordering in time of the distributions presented differed in each of them. Interviewees who marked conflicting answers in questions 11a and 11b (19% in A&B and the Polish sample) often did realize it, but argued that emotions made those situations different:

> "Logically, I should say that it remains the same, (…) but I have a feeling that this inequality will be higher, because of situations between people. If they lived peacefully at the level of 10 or 1000 or whatever and now they will have to count every penny to somehow manage to survive, then the atmosphere in those groups and between them will cause the inequality to be higher" [respondent no. 1406].

Viewing losses and gains differently is a proven phenomenon (Kahneman and Tversky 1979) that often influences survey results. Additionally, respondents noticed that they were asked twice about the same thing and they might have felt tested instead of being asked for an opinion. In this case it seems that asking only one translation invariance question would suffice since the second one does not increase our understanding of the respondents significantly.

Another issue that could have caused some illogical or puzzling answers is the construction of verbal questions. In each of the above questions (Q10, Q11a and Q11b) there were three possible answers: inequality remains the same; inequality has increased and inequality has fallen. However, the order and phrasing of these answers differs in each one. This is especially confusing in case of questions 11a and 11b, since they depict similar situations. In question 10 all of the possible responses have explanations attached to them, in 11a only one, while in 11b none. An argument backing up one of the responses might increase or decrease its popularity (or even influence each respondent differently). For example, one person commented that they chose a different answer than the one they originally intended to in question 10, because the explanation attached to it was an "improper way of thinking" [re. no. 7610]. Amiel and Cowell hoped that by showing respondents that their views can be found in the literature, or in what other respondents suggested, would increase their understanding of the matter at hand. However, they admit (Amiel and Cowell 1992, p. 7) that the influence of these arguments in respondents' answers is unknown and might be insignificant.

Table 3

Income transformation – answers to verbal questions compared

A&C		Add fixed sum (q11a) (%)		
		Down	**Up**	**Same**
Double income (q10)	Down (%)	7	**1**	**4**
	Up (%)	21	2	17**
	Same (%)	30*	**3**	**14**

Source: Amiel and Cowell (1992).

Poland (N = 132)		Add fixed sum (q11a) (%)		
		Down	**Up**	**Same**
Double income (q10)	Down (%)	4	**0**	**11**
	Up (%)	18	2	20**
	Same (%)	16*	**1**	**29**

* translation invariance supporters
** scale invariance supporters
Bold: puzzling answers

Source: own research.

Amiel and Cowell declared their numerical and verbal questions uncomparable (1992, p. 13), but comparing those answers gives us additional information about the respondents' perspective and more accurate information about the level of axiom acceptance. If someone supports an axiom, it should be shown in almost all answers concerning it, not just in the one type of question. It seems though that by taking this approach we are left with very few supporters (eight for scale and ten for translation invariance in the Polish sample[4]) and many intermediate or undecided ones. In the Polish sample answers to numerical and verbal questions about income transformations differed and most interviewees were unaware of it. The questionnaire did encourage participants to reconcile responses to both types of questions and almost half of the sample (47.4%) did use that option. However, not everyone compared their answers to both types of questions and many interviewees reported that they "(…) didn't pay much attention to it" [respondent no. 2088] or found the task tiresome. It is not surprising, since checking ones'

[4] When considering answers to questions 1, 2, 10, 11a and 11b.

answers to the first part of the questionnaire required flipping through the pages back and forth. If all questions concerning the same topic were on the same page it would be easier to see a conflict between responses, however, it would ruin the "learning by doing" approach. Another way of solving this problem without changing the structure of the study would have required creating an interactive computer-based questionnaire that would show the respondent and his previous answers exactly when they were needed.

4.2. Population replication

Population replication states that "cloning" a society, even multiple times, does not change the level of inequality in it. It views people in terms of the percentage of population (e.g. one person is viewed only as 10% of a society of size 10) and is widely accepted by experts. Especially since it is the only axiom that deals with comparing inequality between differently sized groups. Both questions concerning this axiom were about doubling population size.

Q3. A = (5, 8, 10) B = (5, 5, 8, 8, 10, 10)

Q12. Suppose we replicate a three-person society by merging it with an exact copy of itself (so that we now have a society of six people consisting of three sets of identical twins).

(a) The income inequality of the six-person community is the same as that of the three-person community because the relative income shares remain unchanged.

(b) The income inequality of the six-person community is less than that of three-person community because in the six-person community there are some people who have the same income.

(c) The income inequality of the six-person community is greater than that of the three-person community.

Most respondents supported population replication: in the original sample 58% agreed with it in the numerical and 66% in the verbal question, while the corresponding percentages in the Polish sample were 51% and 60% (Table 4)[5]. The second most popular answer was that inequality among the bigger population is lower. Interviewees who gave such a reply pointed out the same thing that appeared in the verbal question: that in the multiplied

[5] When both verbal and numerical questions were taken into consideration the support for population replication in the Polish sample fell to 45%.

group everyone has someone in the same situation as them, so no one is alone or ostracized.

Table 4

Population replication – answers to numerical and verbal questions

Sample	N	Numerical (q3) (%)			Verbal (q12) (%)		
		Down	Up	Same	Down	Up	Same
A&C	1108	31	10	58*	22	9	66*
Poland	131	40	9	51*	31	8	60*

* population replication supporters

Source: Amiel and Cowell (1992), own research.

Questions concerning population replication were rarely a source of conflicting answers. Only one fourth of the Polish sample (26%) replied to them in an inconsistent way. Nevertheless, these discrepancies between verbal and numerical questions touch a very important problem: are the verbal and numerical questions comparable? In the first part of the questionnaire, respondents face concrete examples that are viewed as static situations in different groups or countries. However, the verbal questions depict a situation of an income change in one society: they add a time order of events and that introduces the idea of losses and gains, not just differences. Thus, answers to verbal questions are more emotional, which sometimes cause respondents to answer in a conflicting way:

> "(…) because gaining, and taking away, I don't know, it may have conditioned my change of answers in some situations" [respondent no. 6565].

Another factor that influences answers to verbal questions is the fact that they are verbal: they use certain phrases that may have complex connotations, they trigger emotions, and they use examples, which also creates a framing effect. In the case of population replication we can talk about cloning a society, or about a next, bigger generation with the same incomes, or about merging it with a society that is exactly the same, and so on. The choice of a concrete example might be influential. Also influential is the specific phrasing used. Whether we talk about "rich" and "poor" or "those who have more", or even introduce the idea of "basic needs", it has an impact on the respondents way of thinking. All these factors make verbal questions significantly different from numerical ones. However, this is also the strength of verbal questions,

for they put axioms and problems of income inequality in a setting that is more natural, depicting arguments found in real-life discussions. Verbal questions elicit views on income inequality from a significantly different perspective than numerical ones. Thus, discrepancies between the two are not surprising, yet we should be aware of both of them. Answers to verbal and numerical questions give us partially different data that complement each other and are both necessary to fully understand how people perceive inequality.

4.3. Transfer principle

According to the Pigou-Dalton's principle of transfers, transferring a small amount of money form a richer person to a poorer one decreases inequality. Amiel and Cowell analyse the principle of transfers very closely, they even dedicated a separate questionnaire to it in their later study (Amiel, Cowell and Slottje, 2004).

Q4. A = (1, 4, 7, 10, 13) B = (1, 5, 6, 10, 13)

Q13. Suppose we transfer income from a person who has more income to person who has less, without changing anyone else's income. After the transfer the person who formerly had more still has more.

(a) Income inequality in this society has fallen.

(b) The relative position of others has also changed as a consequence of this transfer. Therefore we cannot say, a priori, how inequality has changed.

(c) Neither of the above.

Amiel and Cowell were surprised by the pattern of acceptance of this axiom. They comment on the results of the numerical question saying that: "nearly two thirds of the sample fail to agree with the transfer principle" (1992, p. 16). When the axiom was presented verbally, 60% of the original sample supported it (Table 5). Results from the Polish sample are very similar: 38% support the axiom in the numerical and 55% in the verbal question. What is more important though is that in the Polish sample answers to verbal and numerical questions seem unrelated[6] (Table 6), as if those questions concerned two completely different topics. This suggests that either those two types of questions show us a completely separate perspective on the problem or, which is more likely, that one or both of them were in some way misinterpreted by the respondents.

[6] If the sample was random, the independence hypothesis would be accepted by Pearson's Chi-Square test at a significance level of 0.05.

Table 5

Principle of transfers – answers to numerical and verbal questions

Sample	N	Numerical (q4) (%)			Verbal (q13) (%)		
		Agree	Strongly disagree	Disagree	(a) Agree	(b) Strongly disagree	(c) Disagree
A&C	1108	35	42	22	60	24	14
Poland	132	38	42	20	55	23	22

Source: Amiel and Cowell (1992), own research.

Table 6

Principle of transfers
– answers to numerical verbal questions compared with the Polish sample

Transfer principle		Verbal (q13) (%)		
		(a) Agree	(b) Strongly disagree	(c) Disagree
Numerical (q4)	Down	23	9	6
	Up	20	9	13
	Same	11	5	3

Note: In the original sample the percentage of those who agreed with the axiom both in numerical and verbal question was 36% (Amiel and Cowell, 1992, p. 17).

Source: own research.

The numerical question concerning the principle of transfers is tricky. Interviewees who disagreed with the transfer principle in the numerical question often pointed out that the first example (1, 4, 7, 10, 13) is very regular in nature because the gaps between consecutive incomes are the same "(…) there is exactly a straight line here (…) and here everything is so perfectly arranged" [respondent no. 1616]. The second distribution on the other hand disrupts this sequence breaking this regularity and leaves the poorest person further away from others. That is why many respondents viewed this change as increasing inequality.

Asking about the transfer principle by describing it straightforwardly yielded more answers agreeing with the axiom, but as Amiel and Cowell noted:

> "those whose responses differed as between question 4 and question 13, the majority did not indicate any desire to go back and change their response to question 4" (1992, p. 17).

The interviews revealed two explanations of this phenomenon. First: most interviewees did not notice that a transfer took place in the numerical question, so they did not take it into consideration. Only those who actually compared their numerical and verbal answers had a chance to even notice the problem. Second: the transfer principle sounds "right". In theory respondents understood that the rule was general, but they rather thought about the most extreme and thus obvious examples of its application. The scenario of a poor person giving to those even poorer did not appeal to them. After being confronted with the conflict in their answers concerning the principle of transfer the interviewees either withdrew their support for the axiom or changed their numerical answer according to it, but they did both these things rather reluctantly. One of the interviewees even said that: "I mean, the principle that moving money to the poorer [people] seems fair, except in the case here (1,5,6,10,13)" [respondent no. 8932]. So she did claim to agree with the axiom, but with this one exception.

Another problem often reported by the interviewees when discussing the transfer principle questions was the way in which numerical examples were shown. Amiel and Cowell made the numerical part of the questionnaire minimalistic, as they said:

> "The distributions are presented as vectors, without explicit currency units, and no hints were provided to the students as to what sort of living standards or welfare levels might correspond to those numbers" (1992, p. 6).

Also, there was no explanation among who the income is distributed: groups, people or working people. The numbers were relatively small ranging from 1 to 20, so they would be interpreted as abstract sums. However, some respondents complained that this resulted in small differences between the examples (only one unit of income was transferred in this example) and difficulties in imagining the situation. Therefore the interviewees made assumptions of their own and some of them even transformed given examples, so that they would be easier to visualise. Depending on a person, the same number, for example 1 meant 1 zloty, 100 zlotys or 1000 zlotys – respondents multiplied those numbers freely to make them easier to imagine. However, the same people did not always agree with the scale invariance axiom (so some of them did reply that multiplication changes inequality). As a result, they answered a different question than the one posed. This is a problem one must be aware of, but for which there is no easy solution. One could express examples in a local currency and use amounts close to incomes achievable in that given society. This would stop respondents from transforming the exam-

ples, but would cause problems with interpreting examples through the lens of the assumed basic needs and respectable income. The usage of abstract sums, instead of real-life earnings, was dictated by the fact that Amiel and Cowell wanted to avoid these aforementioned problems and obtain inter-country comparability of answers.

This situation proves again that asking about axiom acceptance only verbally or only numerically does not give us the full picture. Here both answers contain different kinds of biases and comparing answers to both of them alerts us to existing problems. Therefore, asking both questions does give us additional information and makes data more reliable, but creates the problem of conflicting answers that needs to be somehow solved.

4.4. Decomposability

The decomposability axiom states that inequality in a whole society comes from inequality within groups and differences between them (e.g. a divide between the incomes of men and women). That is why Theil index (which has this property) is often used for in-depth analysis even though its interpretation is quite unintuitive. Three questions in the questionnaire were dedicated to decomposability: two numerical and one verbal. Again we can see serious discrepancies in the axiom's support between verbal and numerical questions, yet this time they were the other way round: the numerical questions suggest bigger support for the axiom than the verbal one.

Q5.	A = (4, 8, 9)	B = (5, 6, 10)
Q6.	A = (4, 7, 7, 8, 9)	B = (5, 6, 7, 7, 10)

Q14. Suppose there are two societies, A and B, with the same number of people and with the same total income, but with different distributions of income. Society A is now merged with C, and society B is merged with C' where C and C' are identical.

(a) The society that had the more unequal income distribution before the merger still has the more unequal distribution after the merger.

(b) We can't say which society has the more unequal income distribution unless we know the exact distributions.

(c) Neither of the above.

Analysing support for decomposability in the numerical part requires two questions. The first one (q5) asks about the difference in inequality between

two small communities: (4, 8, 9) and (5, 6, 10). The second question (q6) compares those two communities both merged with a two-person group (7, 7). An answer is in accordance with the decomposability axiom if it is identical in both questions. Many respondents did that without realizing (at least at first) what the connection between those two questions was. Decomposability is quite a complicated axiom to explain, so the verbal question is also difficult: it requires comparing inequality in four abstract societies. In order to maintain precision, Amiel and Cowell resorted to the use of formal language hence reading the question properly requires concentration and effort. Perhaps using an example e.g. of two countries to which a small region is joined would allow for usage of more natural language and make the question simpler.

Support for decomposability in the A&C sample was quite strong: 57% in the numerical and 40% in the verbal part. In the Polish sample it was 49% and 27% respectively (Table 7), but only 16% expressed support in both of those questions simultaneously. What is more important is that again the pattern of answers suggests a complete lack of relation between those two questions[7]. Most respondents did not notice the way the numerical questions on decomposability were created. Even after reading the verbal questions, it requires a moment of attention to see what exactly happened and what the answers supporting the axiom should look like. Additionally, in each case both distributions are a mirror image of each other, the second example is the result of deducting the first one from 14. However, this is easier to notice in the three-person society than in the five-person one. Some respondents answered "the same" to the first one, because they recognized this, but gave a different answer to the second question, where this symmetry was harder to detect. Using a different, less distinctive example and making the verbal question easier would improve the answer's consistency.

Table 7

Decomposability – answers to numerical and verbal questions

Sample	N	Numerical (q5 & q6) (%)		Verbal (q14) (%)		
		Same	Different	(a) Agree	(b) Strongly disagree	(c) Disagree
A&C	1108	57*	41	40	45	11
Poland	132	51*	49	27	69	5

* decomposability supporters

Source: Amiel and Cowell (1992), own research.

[7] If the sample was random, the independence hypothesis would be accepted by Pearson's Chi-Square test at a significance level of 0.05.

4.5. Unbalanced enrichment

Temkin (1986) described an example of unbalanced immiserisation where an originally perfectly equal and rich society gets poorer and poorer (one person at a time) until everyone is equally poor. Amiel and Cowell decided to analyse how respondents judge such a sequence of events but in a verbal question they assumed reverse order, namely unbalanced enrichment.

Q7.	A = (5, 5, 5, 10)	B = (5, 5, 10, 10)
Q8.	A = (5, 5, 10, 10)	B = (5, 10, 10, 10)
Q9.	A = (5, 5, 5, 10)	B = (5, 10, 10, 10)

Q15. Suppose there is a society consisting of n people. There is one rich person and n-1 identical poor people. One by one, some of those who were poor acquire the same income as the rich person, so that eventually there are n-1 (identical) rich people and just one poor person. Please circle the appropriate response

(a) Inequality increases continuously.
(b) Inequality decreases continuously.
(c) Inequality at first increases and then decreases.
(d) Inequality at first decreases and then increases.
(e) Inequality remains the same throughout.
(f) None of the above.

Interviewees often described questions concerning unbalanced enrichment as "difficult" ("the last three [numerical questions] were, well, a bit hard to answer" [respondent no. 7547]) and giving conflicting answers to them was common. The first type of conflict that occurred in the numerical questions where three income distributions were compared pairwise ((5, 5, 5, 10) (5, 5, 10, 10) (5, 10, 10, 10)). There are 27 possible combinations of answers to those numerical questions and only 13 of them provide a consistent non-conflicting ranking. Over 10% of each sample gave such an illogical response (11% in A&C sample (1999, p. 82) and 12.1% in the Polish sample). Posing this same question in the form of a ranking would reduce error and make it easier to fill in and understand for the respondents.

The second type of conflict, and a much more popular one, was between answers to numerical and verbal questions. The verbal question about unbalanced enrichment had six possible answers, which in itself makes answering hard ("Here, for example, in the 15th [question] there were a lot of options

and one had to read each one very carefully, to be sure" [respondent no. 2088]). Comparing one's verbal answer with the numerical one required a lot of effort and concentration; some interviewees said that even after long consideration they still were not sure if they did it right in the end. Among Polish students who gave a logical response to the numerical questions, 56.1% gave a conflicting answer in the verbal part.

Table 8

Unbalanced enrichment – answers to numerical and verbal questions

Inequality…	Sample (%)			
	A&C		Polish	
	Numerical	Verbal	Numerical	Verbal
Increases continuously	8	7	4	5
Decreases continuously	8	20	12	28
First increases then decreases	26	19	19	16
First decreases then increases	42	35	48	35
Remains the same	3	11	4	11
Does none of the above	1	4	0	5
No transitive answer	11	–	12	–
Partial answer	1	4	2	0
N =	1108		132	

Source: Amiel and Cowell (1999), own research.

Percentages of answers to numerical and verbal questions in the Polish sample are very similar to those received by Amiel and Cowell (Table 8). Their later research showed (1999) the answers to these particular verbal questions were very volatile and are dependant heavily on the way the question is posed. Devooght (2003, pp. 254–256) has also shown that in numerical questions the order in which the distributions are presented (unbalanced enrichment or immiserisation) influences the results significantly. It suggests that the received results are heavily biased. The example and possible answers assume continuity in changes of inequality while the groups get richer one by one. This assumption itself should be verified with the respondents. In the questionnaire, there was no logical way to answer numerical questions

that would show discontinuity since there were only three distributions compared. In the verbal question one could express such a view only by marking "None of the above" which is an answer generally chosen reluctantly. One verbal question and three distributions dedicated to such a meaning loaded example is not enough. The example of unbalanced enrichment comes from a completely different approach to the measurement of inequality than the axiomatic framework and so should have a separate questionnaire dedicated to it.

5. Discussion and conclusion

Research results thus far show a general lack of support for the most common inequality measures and significant differences in responses to verbal and numerical questions. Our study replicated the one conducted by Amiel and Cowell (1992; 1999), adding to prior knowledge new data on a sample of Polish students, and expanded our understanding of the obtained answers through qualitative interviews with selected respondents. Results from the Polish sample were only roughly similar to the ones obtained by Amiel and Cowell. The differences in answers from all the samples could be the result of cultural, language and educational differences, which cannot be overlooked. However, all the samples included both in the original and replicated study are non-random. Thus no statistical test of significance in differences can apply. Furthermore, the results obtained cannot be treated as representative for any bigger population. It would be most beneficial to conduct similar research on a random sample, so that results could be treated as the views of a larger group of people.

Before such a research can be conducted on a larger scale, we should work on further improving and experimenting with the methodology, so that upon receiving data from a bigger random sample, we can interpret them correctly without doubts and questions that are posed by the results obtained so far. Formulation and construction of examples for the numerical part of the questionnaire should be improved in order to minimize errors. The effects of different phrasing of verbal questions should be assessed. Finally, we should search for a way to motivate respondents to connect numerical examples and verbal questions about axioms and for them to think about them jointly. All these issues can be addressed in further studies on small samples.

The research conducted by Amiel and Cowell showed that the way we measure income inequality might be inadequate to the way ordinary people

actually perceive and define it. Their research can be compared to experiments on decision making that showed huge discrepancies between John von Neumann's and Oscar Morgenstern's (1944) rationality axioms and real life human behaviours. These experiments led to a significant development of decision-making theory (a description can be found in Kahneman, 2012). Nevertheless, Amiel and Cowell's study is only a first step, a suggestion that income inequality measurement axioms are not supported by the way people actually perceive inequality. However, their results are undermined by many inconsistencies in their respondents' answers and problems with interpreting others. Furthermore, these results are purely negative for they criticize existing axioms and offer no alternatives. Devising new axioms, or new ways to measure inequality, requires a different methodology and approach. Yet again, before we proceed on the search for a new way to define income inequality, we must first make sure that truly the way we do it now is not the best.

Reassessing, the results obtained by Amiel and Cowell's questionnaire both in the original and in the Polish study are burdened with high data error, but they bring to light important problems and questions. Improvements of research methodology proposed in this paper, based on qualitative interviews and methodological analysis, should enable us to better measure the way "ordinary people" define income inequality.

Bibliography

Amiel Y., Cowell F.A. (1992), *Measurement of income inequality: Experimental test by questionnaire*, „Journal of Public Economics" (47), 3–26.

Amiel Y., Cowell F.A. (1999), *Thinking about Inequality*, Cambridge: Cambridge University Press.

Amiel Y., Cowell F.A. (2001), *Risk and Inequality Perceptions*, „Distributional Analysis Research Programme Papers" (55).

Amiel Y., Cowell F.A. (2002), *Attitudes Towards Risk and Inequality: A Questionnaire-Experimental Approach*, [in:] F. Andersson, H.J. Holm (eds.), *Experimental Economics Financial Markets*, Auctions and Decision Making, (85–115), Kluwer.

Amiel Y., Cowell F.A., Slottje D.J. (2004), *Why Do People Violate the Transfer Principle? Evidence from Educational Sample Surveys*, „Research on Economic Inequality" (11), 1–16.

Amiel Y., Cowell F., Gaertner W. (2012), *Distributional orderings: An approach with seven flavours*, „Theory and Decision" (73), 381–399.

Amiel Y., Cowell F., Ramos X. (2010), *Poles Apart? An Analysis of The Meaning of Polarization*, „Review of Income and Wealth" (56), 23–46.

Ballano C., Ruiz-Castillo U.J. (1993), *Searching by questionnaire for the meaning of income inequality*, „Revista española de economía", 10(2), 233–259.

Cuena E.C., Neugebauer T., Seidl C. (2004), *Leaky bucket Paradoxes in income inequality perceptions: an experimental investigation*, „Economics Working Papers" (04).

Cowell F.A. (1985), *'A Fair Suck of the Sauce Bottle' or What Do You Mean by Inequality?*, „Economic Record", 61(2), 567–579.

Decoster A., Schokkaert E. (2002), *The choice of inequality measure in empirical research on distributive judgements*, „Journal of Economics", 9(1), 197–222.

Davidovitz L., Kroll Y. (1999), *Choices in Egalitarian Distribution Inequality Aversion versus Risk Aversion*, „Distributional Analysis Research Programme Papers" (43).

Devooght K. (2003), *Measuring inequality by counting complaints theory and empirics*, „Economics and Philosophy", 19(2), 241–263.

Harrison E., Seidl C. (1994), *Perceptional inequality and preferential judgements: An empirical examination of distributional axioms*, Public Choice, 79(1–2), 61–81.

Kahneman D. (2012), *Pułapki myślenia szybkiego i wolnego*, Poznań, Media Rodzina.

Kahneman D., Tversky A. (1979), *Prospect theory: An analysis of decision under risk*, „Econometrica", 47(2), 263–291.

Neumann J. von, Morgenstern O. (1944), *Theory of Games and Economic Behavior*, Princeton University Press, Princeton.

Sen A. (1973), *On Economic Inequality*, Oxford, Clarendon Press.

Temkin L.S. (1986), *Inequality*, „Philosophy and Public Affairs", 15(2), 99–121.

Temkin L.S. (1993), *Inequality*, New York, Oxford University Press.

Traub S., Seidl C., Schmidt U. (2003), *Lorenz, Pareto, Pigou: Who scores best? Experimental evidence on dominance relations of income distributions*, „Economics Working Papers" (04).

Traub S., Seidl C., Schmidt U. (2009), *An experimental study on individual choice, social welfare and social preferences*, „European Economic Review", 53(4), 385–400.

Wilkinson R.G., Pickett K. (2010), *The spirit level why greater equality makes societies stronger*, New York, Bloomsbury Press.

MEASUREMENT OF INCOME INEQUALITY RE-EXAMINED: CONSTRUCTING EXPERIMENTAL TESTS BY QUESTIONNAIRE

Summary

Learning ordinary people's perception of income inequality is a relatively new field of research, but its importance is rapidly growing. Despite the popularity of inequality measures, the underlying assumptions concerning the most popular of them do not fully hold, as shown in early research by Amiel and Cowell (1992; 1999). However, results of their study contain multiple puzzling and inconsistent answers. This paper analyses methods used by Amiel and Cowell to elicit respondents' views about income inequality. It presents the results of a quantitative questionnaire repeated after Amiel and Cowell, combined with qualitative interviews with selected respondents. The research was conducted in Poland on 132 sociology and economy students. Qualitative interviews and the subsequent data analysis revealed multiple problems that caused respondents to answer inconsistently. Solutions to some of these problems are then proposed.

POMIAR NIERÓWNOŚCI DOCHODÓW: PONOWNA ANALIZA EKSPERYMENTÓW ANKIETOWYCH

Streszczenie

Próby zrozumienia, jak zwykli ludzie postrzegają nierówność dochodów, to stosunkowo nowa dziedzina badań o szybko rosnącym znaczeniu. Mimo powszechnego stosowania miar nierówności założenia, które leżą u ich podstaw, nie są w pełni akceptowane, co pokazali w swoich badaniach Amiel i Cowell (1992; 1999). Jednakże wyniki tych badań zawierają wiele niespójnych i trudnych do zrozumienia odpowiedzi. Niniejszy artykuł analizuje metodologię zastosowaną przez Amiela i Cowella do poznania poglądów ludzi na temat nierówności w dochodach. Przedstawiono w nim wyniki powtórzenia badań Amiela i Cowella (1992), uzupełnionego o serię wywiadów jakościowych z wybranymi respondentami. Badanie przeprowadzono w Polsce na 132 studentach ekonomii i socjologii. Wywiady jakościowe oraz późniejsza analiza danych ujawniły wiele problemów, które powodowały udzielanie niespójnych odpowiedzi. W artykule zawarto również możliwe rozwiązania zauważonych problemów.

Измерения неравенства доходов: повторный анализ результатов экспериментального анкетирования

Резюме

Попытки выяснения того, как обычные люди воспринимают неравенство доходов, принадлежат к относительно новой области исследований, значение и роль которой возрастают в быстром темпе. Несмотря на широкое применение измерений неравенства, положения, которые приняты за их основу, не признаны до конца, что показали в своих исследованиях Амиель и Коуэлл (1992; 1999). Тем не менее, результаты этих исследований содержат много непоследовательных и сложных для понимания ответов. В настоящей статье представлен анализ методологии, взятой за основу Амиэлем и Коуэллом для ознакомления с мнениями людей относительно неравенства в доходах. Представлены результаты повторного исследования Амиэля и Коуэлла (1992), дополненного рядом качественных методов интервью с избранными респондентами. Исследование было проведено в Польше со 132 студентами факультетов экономии и социологии. Качественные методы исследования, а также последующий анализ данных, выявили множество проблем, которые были вызваны непоследовательностью ответов. В статье предложены также возможные решения обнаруженных проблем.