NBER WORKING PAPER SERIES

ECONOMICS AND MEASUREMENT:
NEW MEASURES TO MODEL DECISION MAKING

Ingvild Almås
Orazio Attanasio
Pamela Jervis

Working Paper 30839
http://www.nber.org/papers/w30839

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2023

This paper is based on the Presidential Address Attanasio presented at the World Congress of the Econometric Society in August 2020. At different stages, we received extremely valuable feedback from many colleagues, including Joe Altonji, Manuel Arellano, Jere Behrman, Alberto Bisin, Richard Blundell, Margheria Borella, Agar Brugiavini, Andrew Caplin, Rossella Calvi, Sarah Cattan, Flavio Cunha, Maria Cristina De Nardi, Aureo de Paula, Armin Falk, Jim Heckman, Mike Keane, Sonya Krutikova, Chuck Manski, Costas Meghir, Rohini Pande, Elena Pastorino, Larry Samuelson, GuglielmoWeber, Ken Wolpin. Attanasio presented some of this material in seminars at the University of Chicago, Bonn University, the European University Institute, St Andrews. We thank Marianne Moreira and Diana Lopez Perez for capable research assistance. We are grateful to the British Academy and the European Research Center for the provision of grants that helped the collection of data in Tanzania, including Attanasio's ERC Advanced Gran 695300 (HKADeC). Jervis gratefully acknowledges financial support from the Institute for Research in Market Imperfections and Public Policy MIPP (ICS13 002 ANID) and the Center for Research in Inclusive Education, Chile (SCIA ANID CIE160009). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Economics and Measurement: New measures to model decision making
Ingvild Almås, Orazio Attanasio, and Pamela Jervis
NBER Working Paper No. 30839
January 2023
JEL No. A1,D1

ABSTRACT

Most empirical work in economics has considered only a narrow set of measures as meaningful and useful to characterize individual behavior, a restriction justified by the difficulties in collecting a wider set. However, this approach often forces the use of strong assumptions to estimate the parameters that inform individual behavior and identify causal links. In this paper, we argue that a more flexible and broader approach to measurement could be extremely useful and allow the estimation of richer and more realistic models that rest on weaker identifying assumptions. We argue that the design of measurement tools should interact with, and depend on, the models economists use. Measurement is not a substitute for rigorous theory, it is an important complement to it, and should be developed in parallel to it. We illustrate these arguments with a model of parental behavior estimated on pilot data that combines conventional measures with novel ones.

Ingvild Almås
Institute for International Economic Studies
Stockholm University
SE-106 91 Stockholm
Sweden
Ingvild.Almas@nhh.no

Orazio Attanasio
Department of Economics
Yale University
37 Hillhouse Avenue
New Haven, CT 06511
and Institute for Fiscal Studies,
FAIR, BREAD and CEPR
and also NBER
orazio.attanasio@yale.edu

Pamela Jervis
Department of Industrial Engineering
University of Chile
Beauchef 851
Santiago, Chile
and Institute for Fiscal Studies
pjervisr@uchile.cl

1 Introduction

For many years, at least since Samuelson's (1938; 1948) and Arrow's (1959) contributions, most economists thought that good empirical work can only be based on a narrow set of measures. The prevalent perception, with a few exceptions, was that we can only meaningfully use measures of what people buy or do, their resources, prices, and (possibly) the markets they have access to. While many economists have also proficiently used *objective* data, including biological and anthropometric data, the set of measures deemed acceptable and interesting by a large fraction of the profession has been limited. Data on stated preferences and their intensity, intentions to buy, stated actions in hypothetical situations, subjective expectations, and attitudes and tastes – as well as data on the possible drivers of individual choices, such as social norms, culture, or political attitudes – were seen as not particularly useful and outside the realm of economics.

From the measures perceived as meaningful and acceptable one could, under certain assumptions, infer and characterize preferences and other structural parameters that drive individual behavior, as well as some of the features of market structure and, maybe, identify empirically the causal links among different variables. To perform such exercises on a restricted set of measures, however, requires strong assumptions.

The reliance on restrictive and specific sets of measures, disregarding richer measures, such as data on rankings of different choices or the intensity of preference, data on stated intentions, or answers to questions about choices in hypothetical scenarios has been widespread in different economic fields and types of models, both static and dynamic. Such an approach has forced the use of restrictive models in empirical analysis.

In most static models, for instance, empirical research has used mostly choice-based data in combination with objectively observed variables (such as prices or other individual and environmental data).[1] This approach has led researchers to focus on models that either imposed homogeneity assumptions or a very specific set of heterogeneous preferences. And even in models with heterogeneous preferences, such as the Random Utility Model, the use of choice-based data poses key identification problems. Such problems are particularly salient when individual choices are determined not only by preferences and resources but also by other factors that are typically deemed unobserv-

[1] Berry et al. (2004) is an exception, where rankings or 'second hypothetical choices' are used in the identification of demand models for cars.

able, such as individual beliefs.[2] In the case of dynamic models, where uncertain future variables play a role, researchers used assumptions such as adaptive, myopic or, in recent decades, rational expectations, which allowed an internally consistent solution of the model under study and its empirical applications.

The kind of measures researchers consider as meaningful and usable implicitly defines the models one can bring to data and, therefore, the domain of economics as a social science with empirical content. In this paper, we discuss the important role new measures can play in the study of economic behavior. The use of innovative measures, such as answers to hypothetical questions used in conjunction with choice data, can help the identification of causal links with less restrictive assumptions. It also limits the need to identify exogenous sources of variation in observational data.

We discuss what should be measured and how these measures relate to models of economic behavior. While we stress that rigorous and coherent models of human behavior and human interactions are key to understand economic reality, we do not think such an approach implies that empirical studies should exclusively use measures derived from choice data and easily and objectively measurable quantities (such as prices or anthropometric data). Instead, we argue that economic theory can and should inform the design of new measures that capture the factors salient for the models at hand. These measures can lead to the use of more flexible and richer models of individual behavior.

One of the main reasons a large part of the profession shied away from certain types of measures has been the challenge in designing appropriate tools to gather them. Therefore, a substantial amount of research effort should be devoted to the design and, importantly, the *validation* of new measurement tools, to ensure that they capture the phenomena they are meant to. New measures, designed and devised by researchers, should be piloted, validated and shown to be correlated with individual choices. Appropriate techniques in psychology and survey design can (and should) be used to develop tools that could capture latent variables relevant for models of individual behavior.

Measurement research and progress are not new. An example is the development of the unfolding bracket technique (Juster et al., 2006), to measure variables, such as household wealth, which had been seen as impossible to assess accurately. New mea-

[2] A very recent paper by Dardanoni et al. (2022) discusses the restrictions that the use of choice data requires for models with preference and other types of heterogeneity.

sures continue to emerge: Bloom and Van Reenen (2007), for instance, started a new research agenda to construct innovative measures of *management skills*, seen as inputs of a production function. Caplin (2021) has been proposing the design of new measures thorough *data engineering* with a logic close to what we argue for in this paper.

A focus on measurement does not remove the need for theoretical models. Indeed, the use and design of new measures should be driven by theory and the need to identify key parameters of the theoretical models under study, using direct measures instead of restrictive (structural) assumptions. Further, an approach to measurement and empirical strategies that goes beyond standard measures does not imply rejecting the modeling of individual behavior based on some sort of constrained optimization that satisfies a set of axioms. As we argue below, the use of new and innovative measures allows researchers to give empirical content to flexible and powerful models, which can replace restrictive models constrained by the lack of available measures. The development of new measures and new more flexible models should go hand in hand.

The rest of the paper is organized as follows. In Section 2, we provide some background material about measurement, its role and its interactions with economic theory. In Section 3, we discuss various aspects of the relationship between theory and measurement, and present the case for going beyond standard measures. We discuss what measures are useful for different models, how the parameters of measurement systems can be identified and how they can be relevant for defining the metric of the relevant latent factors and how they can help the identification of causal links between the various latent factors. Having discussed these conceptual issues, we sketch, in Section 4, a model of household behavior where some of the issues on the relationship between theory and measurement are fleshed out and made explicit in terms of the latent factors and drivers that populate the model. In Section 5, we provide some examples of measurement work and its use, using a novel data set that was collected to pilot new measures of the abstract concepts that are used in the model presented in Section 4. In Section 6, we use the new measures and suggest a theory-funded empirical model of parental investment that shows how these can be used in combination with standard ones. Section 7 concludes the paper.

2 The context

A stark statement of a restrictive approach to the study of preferences in economics is in Stigler and Becker (1977), entitled *De Gustibus Non Est Disputandum*:

> "... tastes neither change capriciously nor differ importantly between people. [...] one does not argue over tastes for the same reason that one does not argue over the Rocky Mountains - *both are there, will be there next year, too, and are the same to all men*." (emphasis added).

It is interesting to note the explicit reference to the stationarity of preference and its cross-sectional homogeneity. Although empirical studies based on choice data and objective measures do not necessarily require cross-sectional homogeneity – and indeed much empirical and econometric research has studied models which allow for heterogeneity – this statement implicitly asserts what are acceptable measures for economists and imposes important restrictions on the models researchers can eventually study.

Imposing the use of a limited set of measures implies that *identification* (that is, the possibility of retrieving empirically the fundamental features of the model under study) may only be achieved with strong assumptions on tastes, beliefs, expectations, and information individuals have access to. This is the price that is paid to assume that such variables and factors cannot be observed or measured. However, the profession has shown much skepticism towards novel measures that could provide information on these types of variables, such as questions that pose *hypothetical* situations and evidence from *stated* rather than *actual* choices.

These issues have been analyzed for a long time. An example of the arguments about what and how to measure can be found in the discussions of stated preferences and conjoint analysis, for instance, in Luce (1956, 1959); Luce and Tukey (1964), and Luce and Suppes (1965) Likewise, the discussion of the Random Utility Model by Block and Marschak (1960) states that the way of defining the class of basic observations and testable conditions is to some extent arbitrary and dependent on the range of possible experiments and observations. They further argue that it may be beneficial to follow the practice in psychology of accepting subjects' ranking of objects and intensity of preferences, even if observed through a verbal statement (see also Caplin, 2021, for a discussion of these issues).

There are good reasons for the profession's skepticism about certain measures. Measuring hypothetical choices, preferences, and attitudes is fraught with difficulties. Framing effects, for instance, seem to be pervasive and introduce a number of potentially severe biases. Several studies, such as List and Gallet (2001) and Murphy et al. (2005), discuss common biases in answers to questions about hypothetical situations. An interesting debate in this respect regards the use of *contingent valuation*. While this type of measure is widely used in other disciplines,such as marketing,[3] its use in economics has received a considerable amount of resistance. Hausman (1994) expressed doubts about its usefulness and Hausman (2012) labels the enterprise as hopeless. This skepticism is partly due to measurement difficulties, though it is also probably due to the ambiguities around what one is measuring when asking questions about hypothetical choices.

Other interesting early discussions of what could and should be measured was the lively exchanges between Tobin (1959) and Katona (1959) on the usefulness of data on buying intentions; Tobin strongly criticized the usefulness of such data on the basis that they were not a good predictor of actual consumer choices. The reliability and predictive power of buying intentions and purchasing probabilities data were later discussed in Juster (1964, 1966) and then again in Manski (1990), who noticed that intention questions are not necessarily useless if formulated properly. Manski argues the issue is not *what* is being measured, but the *specific tools and questionnaires* being used.[4]

While these issues were being debated, some researchers used stated preferences and elicitation of hypothetical choices to estimate key parameters of economic models. Juster and Shay (1964), for instance, used the elicitation of stated choices in hypothetical situations to estimate the elasticity of the demand for loans to interest rates and loan maturities; the interest rate and maturity of these *hypothetical loans* were exogenously varied across respondents to the survey. Cross-sectional differences in loan demand elasticities were then used to discuss the importance of liquidity constraints. More recently, Lancaster and Chesher (1983) used, in conjunction with a model of employment search behavior, "the answers to two simple questions" which could be interpreted as providing information about the distribution of offer wages and reservation wages to "...*deduce* structural parameters rather than estimate them" (p. 1661). What could and

[3]Good references are Louviere et al. (2000) and Carson (2012).

[4]Manski also remarked that intention data, while not been used much by economists, are instead widely used in other disciplines, including marketing. Curtin (2016) provides a nice survey.

should be measured and its relation to black theory was discussed in Haavelmo's (1958) Presidential Address to the Econometric Society. Haavelmo perceived measurement questions to be central to the development of economic theory:

> "I think most of us feel that if we could use *explicitly* such variables as, e.g., what people *think* prices or incomes are going to be, or variables expressing what people *think* the effects of their actions are going to be, we would be able to establish relations that could be more accurate and have more explanatory value. *But because the statistics on such variables are not very far developed, we do not take the formulation of theories in terms of these variables seriously enough.* It is my belief that if we can develop more explicit and *a priori* convincing economic models in terms of these variables, which are realities in the minds of people even if they are not in the current statistical yearbooks, then *ways and means can and will eventually be found to obtain actual measurements of such data.*" (emph. added).

Haavelmo's address argues new measurement tools should be developed in reaction to theoretical models that are both rich enough to capture complex phenomena and require novel measurement tools. Such new measurement tools would allow the identification and characterization of richer models, where strong assumptions are relaxed. In what follows, we similarly argue that new measures for variables that play an important role in several theoretical models should be developed and validated.

These early efforts and discussions were not without consequence. For instance, Katona's work (see e.g., Katona, 1974) led to the establishment of the Michigan survey of consumer sentiment, which is still running (and used) today. In recent decades, however, the design and use of such surveys has become rare. From both a theoretical and measurement point of view, the consensus within most of the economic profession went towards an almost exclusive use of choice based and objectively observable data.

An exception is the field of experimental economics, where researchers construct controlled – and often artificial – settings in which subjects make choices and, in so doing, reveal their preferences, beliefs, and other key drivers of behavior (Plott and Smith, 2008). This approach allows researchers to control the situations presented to subjects, typically in a laboratory setting, so to control away many potential confounding factors to reveal features of human decision-making. A considerable amount of such work has been done on mechanisms and protocols to elicit primitives of individual behavior.

However, lab experiments often require participants to behave abstracting from their present circumstances, imposing a separation between experimental and actual behavior. Indeed, background information on experiment participants is rarely collected and experimental data are rarely used in conjunction with observational data.

The expanded use in the field of techniques and protocols developed in labs and the simultaneous collection of experimental and standard observational data are a sign that the economic profession has been changing its approach to what can and should be measured. Evidence on measures of preference and attitudes towards redistribution, attitudes towards migrants, bargaining and social preferences, reciprocity in conflict areas, and willingness to compete through experiments combined with observational data are reported in Almås et al. (2020b), Alesina et al. (2018b), Almås et al. (2018); Buser et al. (2014); Cavatorta and Groom (2020). At the same time, techniques used in empirical studies on standard data have been reproduced in the laboratory, in particular, for the analysis of auction models.[5]

Several other measurement novelties have proliferated in recent years. One important contribution, which has enlarged the set of variables economists consider measurable, is the study of subjective expectations (e.g., about income, inflation, or rates of return) promoted in an important contribution by Manski (2004).

Subjective expectations data were already contained in the early NLS data, sometimes known as the *Parnes data* and discussed in Parnes (1975).[6] Another example of early collection of subjective expectations is Visco (1984). References relevant for the design of effective expectations measures include Dominitz and Manski (1997), Dominitz and Manski (1996), Potter et al. (2017).

The collection of expectations data has evolved considerably and these types of data are now routinely collected. the Bank of Italy has collected subjective expectations data for a number of years and the Federal Reserve Bank of New York has been and similar data are also collected by the Bank of Spain and the European Central Bank. An important example of high quality subjective expectations data are those collected systematically since 2013 by the Federal Reserve Bank of New York: the *Survey of Consumer Expectations* (https://www.newyorkfed.org/microeconomics/sce/background.html).

[5] See Ertaç et al. (2011); Salz and Vespa (2020). We thank Aureo de Paula for pointing this out.

[6] We are grateful to Ken Wolpin for pointing us to these data.

Research economists have also started to use subjective expectations data within structural models of economic behavior. An early use is Wolpin and Gonul (1985), while a non-exhaustive list of more recent studies include Jappelli and Pistaferri (2000), Pistaferri (2001), Pistaferri (2003), Van der Klaauw and Wolpin (2008), Kaufmann and Pistaferri (2009), Attanasio and Augsburg (2016), Paiella and Pistaferri (2017), Attanasio et al. (2018), and Giustinelli et al. (2019).

The availability of subjective expectations data allows researchers to avoid strong assumptions, such as rational expectations. Moreover, such data allow the identification of genuine measures of uncertainty, which, using data on actual realizations of the variable of interest is not easily disentangled from individual heterogeneity or variability that is known and deterministic to individuals. Moreover, the absence of expectations data implies strong limitations to the identification of certain parameters. Chen et al. (2020) show that, without assuming rational expectations (or data on subjective expectations), only set (rather than point) identification can be achieved.

A related topic is the measurement of beliefs about the return to specific investments, such as different types of investment in human capital and education. Rather than assuming that individuals have rational expectations about the returns to certain investments, researchers have started eliciting data on *beliefs* about returns. Several studies have started collecting and analyzing data on parental beliefs, following a practice that has been used for some time in psychology and child development, as surveyed, for instance, by Miller (1988). Examples of such studies include Cunha et al. (2013), Boneva and Rauh (2018), Attanasio et al. (2019b), Attanasio et al. (2019a), and Biroli et al. (2022). Likewise, Dominitz and Manski (1996), Wiswall and Zafar (2015), and Delavande and Zafar (2019) have studied beliefs about the returns to college and college enrollment choices, Bobba and Frisancho (2020) assessed how college application choices are affected by students' perceptions of their own ability and how these can be changed by additional information, while Dizon-Ross (2019) studied how parental beliefs about their children's abilities affect their choices.

Another example of new measurement tools being developed and used by economists is the study of stated preferences and answers to hypothetical questions. A good example of such a practice is Ameriks et al. (2020), which uses a combination of stated preferences and actual choice to identify complex structural models. Bernheim et al.

(2021) discuss how to use data on hypothetical choices to identify causal links in economic models. These approaches are analogous to questions about *intentions* (e.g., how respondents would allocate hypothetical resources among different potential uses), to elicit information about individual tastes and preferences.[7]

While resources, preferences and tastes are obvious drivers of individual choices, other factors can also be important drivers of behavior. In certain contexts, for example, the quantity and quality of information available to individuals is important. Individuals' access to markets or networks can influence the allocation of resources across time and space. Additional factors, such as learning, risk sharing arrangements, preferences about different policy options, attitudes and social norms, might also affect individual choices (and preferences). When studying household-level choices, for instance, who controls resources and bargaining power within the household might be important.

A variety of studies have used new and innovative measures to analyze many of these phenomena. While the scope of this paper is not to provide an exhaustive survey of the relevant literature, we mention Attanasio and Krutikova (2020) on measuring the quality of information in networks and the role that it plays in providing informal insurance, Almås et al. (2018) and Jayachandran et al. (2021) on measuring bargaining power within couples, and Alesina and Angeletos (2005), Almås et al. (2020b), Alesina and La Ferrara (2005), Kuziemko et al. (2015), and Alesina et al. (2018a) on measuring attitudes towards and perceptions of immigration, social mobility, redistribution, and other policy factors. These studies and others use novel measures of attitudes, sometimes in combination with standard survey data, to quantify the effects of such variables.[8] Kaiser and Oswald (2022) uses longitudinal data from three countries to document the predictive value of 'happiness scales' on individuals' important decisions.[9]

These studies illustrate the active and ongoing discussions in economics about measurement and its relation to theory. Recent developments indicate that the profession is moving towards using choice data and directly observable variables in combination with *stated preferences* and answers to hypothetical questions. An interesting and re-

[7]Studies of stated preferences and answers to hypothetical questions include Blass et al. (2010), Kesternich et al. (2013), Ben-Akiva et al. (2019), Harris and Keane (1998), and Erdem et al. (2005).

[8]These issues are discussed extensively by La Ferrara (2019).

[9]This study stresses that comparability across different contexts implies the need to establish a cardinal metric for measures that are often obtained as ordinal indexes, an issue we discuss below.

cent take on measurement and its relation to theory is discussed in Stantcheva (2022): "Surveys are not merely a research tool. They are also not only a way of collecting data. Instead, they involve creating the process that will generate the data." We share this view. In what follows, we develop it to include the design and validation of new measures to be used in combination with traditional ones for the empirical study and characterization of economic models.

3 Measurement and Theory

The *economic models* economists work with often use abstract constructs that are not directly observable. The measures relating to these constructs that economists collect and use are (and should be) driven by the theoretical framework that organizes our thinking. Examples abound in several fields of economics including the work by Keynes, Stone, and others that led to the development of National Accounts and the analysis of consumer demand, which is behind most methods to construct price indices. The measurement of growth, consumption, and price indices has used theoretical models as a foundation for the construction of new measures. Analogously, the creative use of micro-data has informed the development and calibration of theoretical models.[10]

We discuss the relation between theory and measurement in Section 3.1, and, in Section 3.2, we argue that new measurement tools, beyond standard measures, are desirable and useful. In Section 3.3, we discuss some issues relevant for *measurement systems*.

3.1 Economic Models, Latent Factors and Measurement Systems

Many theoretical models in economics can be represented by:

$$F(\theta;\phi) = 0 \tag{1}$$

[10]Classic studies that led to the development of Modern National Accounts include Keynes (1936), Kuznets et al. (1937), Kuznets (1941), Gilbert et al. (1949), and Stone (1984). Examples on the effects of new products and quality on the measurement of inflation include Bils and Klenow (2001), Bils (2009), and Crawford and Neary (2021). For price indices, see Stone (1954), Christensen et al. (1975), Deaton and Muellbauer (1980), and more recently Nordhaus (1998). On the importance of *creative destruction* to measure growth Aghion et al. (2019); and Neary (2004) and Almås (2012) on international comparisons.

where θ is a matrix of variables of interest or factors, some of which are latent, in that they are not necessarily observed. The parameter vector ϕ characterizes the function F, which represents individual behavior and interactions (such as markets), that is, the relevant economic model. F typically defines what the variables of interest are.

Within this framework it is easy to introduce a number of details about the features of the economy under study. For instance, one could include in model (1) uncertainty and imperfect information, and consider additional factors relating to the information available to the model's individuals. The dimension of the model's latent factors depends on the specific issues under study. Richer and more realistic F functions require a richer set of factors and, to be characterized empirically, a richer set of measures.

The *factors* that populate a theoretical model are often well-defined but unobserved variables, such as prices of very finely defined goods or the quality of family environment and schools. In practice, what is often available are *markers* corresponding to (some of) these theoretical constructs. To bring the theoretical models to data and to identify and estimate the parameters ϕ that define the causal links one is interested in, it is necessary to be explicit both about the theoretical model and about the relationship between the relevant latent factors and the available measures. In other words, to give empirical content to the function F in model (1), one needs a *measurement system* that relates the *latent factors in F* to the available measures. A possible mapping is the following:

$$\mathbf{m} = g(\theta, \varepsilon) \tag{2}$$

where $\mathbf{m}$ are available measures related to the (potentially unobservable) factors θ through the function g. The vector ε is measurement error that, together with the possibility t the function g is not injective, prevents the direct observability of (some of the) θ. The model F defines the factors of interest, and, in turn, guides what measures to look for. The available measures and the measurement system in (2) define what latent factors one can study empirically and which models can be taken to data.

Such mapping between latent factors and constructs of interest for economic theory and a set of available measures resonates with Goldberger's (1972) description of the interplay of theory and measurement in his Fischer-Schulz lecture (*emphasis added*):

> "By structural equation models, I refer to stochastic models in which each equation represents a causal link, rather than a mere empirical association. The models arise

in non-experimental situations and are characterized by simultaneity and/or errors in the variables. The errors in the variables may be due to measurement error in the narrow sense, or *to the fact that measurable quantities are not the same as the relevant theoretical quantities*. Generally speaking the structural parameters do not coincide with coefficients of regressions among observable variables, but the model does impose constraints on those regression coefficients".

Goldberger (1972) uses the Permanent Income model as an example, where *permanent income* is the interesting construct and the empirical measures potentially related to it are *current income and consumption*. In a similar vein, Griliches (1974) discusses the relationship between earnings, schooling, and ability. More recently, Cunha et al. (2010) use a relatively flexible version of measurement system (2) to estimate the production function of human capital. In a different context, the estimation of a production function with endogenous inputs could be viewed in a similar fashion.[11] The early work on Multiple Indicators Multiple Causes (MIMIC) models and, more generally, studies on factor models in economics, psychology, sociology, and geneticsare relevant and important.[12]

As, economists' theoretical models are often populated with abstract constructs, recognizing it explicitly is useful. It clarifies the research objectives and may motivate attempts to measure additional relevant variables, which, in turn, can motivate researchers to use more realistic models that are subject to less stringent assumptions.

Which latent factors to measure. The measurement system and the measurement tools used in empirical studies – as well as what should be measured (and possibly how) – should be informed by the specific questions researchers ask and by the theoretical models being used. Expanding the set of objects one measures allows the consideration of more flexible models and avoids strong and sometimes misleading assumptions.

The latent factors of interest depend on the complexity of the theoretical model being studied, which, in turn, might depend on the phenomenon being interpreted. In this regard, an explicit discussion of the restrictions imposed on the theory by data availabil-

[11] See, for instance, Olley and Pakes (1992); Levinsohn and Petrin (2003); Ackerberg et al. (2015); Gandhi et al. (2020); Doraszelski and Jaumandreu (2013), and Doraszelski and Jaumandreu (2018).

[12] See Wright (1934), Duncan (1966), Goldberger (1971), Goldberger (1972), Griliches (1974), Jöreskog and Goldberger (1975), and Chamberlain and Griliches (1975).

ity and measurement challenges is useful. In certain contexts, it might be apparent that these restrictions have negligible impacts for what is being studied; in other contexts, however, restrictive definitions of the relevant latent factors might substantially limit the ability of a model to explain observed phenomena.

When working with demand systems, researchers typically aggregate different commodities in coarse categories. However, available data (even when very detailed) might miss important components of the commodities considered, such as quality of certain commodities or the market structure faced by consumers.[13] When studying intrahousehold allocation, measures of individual-level consumption (in addition household-level expenditure) are often unavailable, forcing strong assumptions.

Consider, for instance, studies of production functions where output is the result of combining different inputs, such as human and physical capital. Until the late 1990s, studies of labor market inequality used a basic model where production is performed via a production function that uses two types of labor (skilled and unskilled), fitted well to a set of labor market facts.[14] That model, however, could not explain what happened in the first part of the 21st century. As a result, new models that disentangle *skills* from *tasks* have been developed, such as in Acemoglu and Autor (2011) and Deming (2017). The empirical needs of these models require new types of data, such as the O*Net or DOT data sets used, for instance, by Autor and Dorn (2013) and Acemoglu and Restrepo (2019). To study empirically more complex production functions with flexible roles of different skills, such as *sociability*, *drive*, and *motivation*, in addition to *cognition*, it is key to measure these skills, how they differ across individuals, how different occupations might require different combinations, and how they are remunerated in the labor market. Likewise, the measurement of these latent factors, and the comparability of measurement tools, often used in different contexts, then becomes particularly challenging and key to the results one obtains. Similarly, in studying firms heterogeneity, Bloom and Van Reenen (2007), Bloom et al. (2019), and Scur et al. (2021) use an in-

[13]Researchers typically use *price indices* for the aggregate measures of commodities that are used in analysis. However, in some contexts prices may not be linear and change with the quantity purchased, as in many models of price discrimination (e.g., Maskin and Riley (1984), Jullien (2000), and Attanasio and Pastorino (2020)). Recent progress has been made with the analysis of scanner data (e.g., Einav et al. (2008), Griffith and O'Connell (2009), and Dubois et al. (2020)).

[14]As discussed, for instance, by Katz and Murphy (1992), which developed Tinbergen's original explanation of labor market inequality as the effect of the relative demand and supply of different skills.

novative survey, now deployed in several countries, to measure managerial skills as an input in the production function.

In summary, the questions a researcher is addressing, the theoretical models they use, and their empirical performance define the key latent factors of interest. This, in turn, defines which measures are needed to give empirical content and bite to the theoretical framework at hand. While in some cases standard measures, possibly anchored by choice data, are sufficient, in many other cases they are not.

3.2 Beyond standard measures

The revealed preference approach has been very useful for structuring our thinking around agents' decision-making and relating rigorous theory to data with relatively few assumptions. With observations about individual choices, the conditions under which these choices are made (such as the resources individuals control and the prices they face), and objectively measurable variables, preferences consistent with a set of axioms which defines a theory of economic behavior may be revealed. Identification of structural parameters can then be achieved *without explicit measures of tastes, beliefs, or attitudes, and without questions about behavior in hypothetical circumstances.*

However, this approach often rests on strong assumptions. Many interesting economic models routinely deviate from such assumptions and constructs, posing difficult questions if their empirical analysis relies exclusively on *choice data and variables objectively measured.* Often these challenges are tackled with strong theoretical or empirical assumptions. For instance, to study dynamic problems where future and uncertain variables are key, researchers frequently have used the assumption of rational expectations and of an efficient use of all the available information.

On the empirical side, to avoid endogeneity issues, researchers try to isolate or create sources of exogenous variation through 'natural' experiments, a strategy that often requires ad-hoc and un-testable assumptions. Even genuinely exogenous variation induced by randomization can often identify a narrowly defined set of parameters.

Expanding the set of measures one uses might make it possible to study more realistic yet still rigorous models. With new and more comprehensive measures, it is possible to maintain consistency with economic theory while at the same time relax some strong assumptions. For instance, questions about behavior in hypothetical scenarios can introduce exogenous variation in a much richer way than natural experiments. While important contributions on new measures have been made in the literature – e.g., the aforementioned literature on subjective expectations – there is still a need to develop and validate new measures that can be used in the analysis of agents' decision-making.

The choices between alternative theoretical approaches should not be forced and limited by the availability (or lack thereof) of data and appropriate measures. For instance, when trying to explain the presence of what has been called *present bias*, an interesting debate is whether one should model time consistent decision makers or allow for the simultaneous presence of present and future selves. While some experimental measures could also be (and have been) devised to measure the presence of present bias, the choice between different models is a theoretical one. Whether present bias is introduced via temptation preferences, as suggested by Gul and Pesendorfer (2011) – which satisfies time consistency and the axiomatic approach, where the decision unit is the same over time – or whether one allows the presence of multiple selves for an individual, it is a theoretical choice. However, the availability of appropriate measures could crucially help researchers discriminate between these alternative models.

The need for and use of new measures: a few examples. In order to clarify the need for new measures that go beyond data on choices and to better understand the framework in which such new measures can be used, this section provides a few examples, some of which relate directly to the applications we present in Sections 4 and 5. These examples are relevant for: (i) the definition of decision units in models of choice; (ii) the separate identification and characterisation of preferences; and (iii) the characterization of the economic environment.

Decision units. A first step when modeling individual behavior is to *define* the decision unit. In standard consumer theory, the household has most often been considered as the relevant decision unit. In such a *unitary* model, the household as a whole is considered the relevant decision unit with well-defined preferences. In recent decades, however, researchers have focused on how resources are controlled and allocated *within* the household, developing alternative models where multiple decision makers, each with distinct preferences, interact within the household to arrive at household-level choices.

The collective model, first proposed in Chiappori (1988), is one such attractive alternative. Its key assumption is that choices, while resulting from interaction between decision-makers with potentially distinct preferences, are efficient. While a number of important theoretical results, have been derived,[15] characterizing the parameters of the models used and testing their validity exclusively with *choice data* on household consumption and expenditure is challenging.

Much more can be learned by using additional information on *private* consumption. A number of researchers have used information at the individual level within households, such as Dercon and Krishnan (2000), Dunbar et al. (2013); and more recently Lechene et al. (2022). However, even when *individual level* data are available, identifying the determinants of individual and eventually household behavior can only be achieved with strong assumptions. For instance, it is difficult to allow for *caring*-preferences, i.e., when one partner cares about the consumption of the other partner. Instead, nonstandard measures that do not rely on the exclusive use of choice-based data can generate important insights about the process of intrahousehold allocation of resources. For instance, hypothetical choice scenarios elicited separately from household members can generate direct information on individual *tastes*. Likewise, it may be possible to derive information on the relative bargaining power within the household, a measure that goes beyond standard choice-based data. We discuss some of these measures in Section 5.1.

Disentangling beliefs and tastes. Agents in most models in economics make decisions to maximize an objective function, given the resources available to them. These decisions then depend on their preferences and on their perception of *the process that links actions to outcomes*. Often, the characterization of such a process is of key interest to researchers. Identifying the causal links that define it requires understanding how individual choices are made. It is often assumed that individual decision-makers *know* the process that determines the outcomes they care about, given their actions and other variables. However, in many situations, this assumption is a strong one.

[15] See, for instance, Browning and Chiappori (1998); Bourguignon et al. (2009); and Cherchye et al. (2011). Attanasio and Lechene (2014) present tests of the collective model using the variability induced in a household demand curve by two different *distribution factors*, i.e. variables that affect Pareto weights but not utility.

The challenges related to disentangling individual perceptions or beliefs and tastes have been extensively discussed in several different settings (see e.g., Caplin, 2021). Possible approaches are the direct elicitation of beliefs (and retrieving preferences from choice data); and the elicitation of preferences through experimental approaches and hypothetical choice scenarios, holding beliefs constant by giving surveys' respondents full information about the context.

Beliefs elicitation. Direct elicitation of beliefs has been used in a model of child development and parental investment in Cunha et al. (2013) and by Attanasio et al. (2019b), eliciting the perceived productivity of parental investment on child development. We follow a similar approach in Section 5. In a different context, Mueller and Spinnewijn (2021), studying on search behavior among unemployed, *suggest using direct measures of beliefs while retrieving tastes as a residual.*

Preferences elicitation. Another approach is to elicit preferences directly *holding beliefs constant* in controlled situations with full information about the actual setting. This can be done either with experimental methods, revealing preferences through real choices, or posing different hypothetical scenarios to respondents and eliciting their (stated) preferences. An early example of such a strategy is the aforementioned Juster and Shay (1964). More recently, Ameriks et al. (2020) and Bernheim et al. (2021) have used hypothetical questions to estimate parameters that characterize individual preferences.

Elicitation of preference and beliefs. Instead of eliciting beliefs (preferences) and inferring preferences (beliefs) as a residual by combining the elicitation with choice data, one may directly elicit both beliefs and preferences in particular choice situations. For instance, in a recent paper Adams and Andrew (2019) who use *survey experiments* to elicit *average* beliefs and preferences in the Indian (specifically Rajasthani) marriage market for young brides.[16] To allow for heterogeneity in both beliefs and preferences it may be beneficial to combine the strength of economic experiments or hypothetical choices to elicit preferences, and the direct elicitation of beliefs through surveys.

[16]Other recent studies include List et al. (2021) who have elicited beliefs from parents in Chicago; Bobba and Frisancho (2020) collect and use data on self-perceptions of academic achievement among high school students in Mexico; Miller et al. (2020) study beliefs about contraception effectiveness.

The environment. To better understand individual (economic) behavior, it is useful to measure how wider constructs, such as institutions, communities, and society at large affect individual behavior via social norms, attitudes, or culture, and model their evolution (see e.g., Bisin and Verdier (2000) for a discussion of how culture evolves).

Social norms and attitudes affect individuals' objective functions in significant ways. In a recent paper, for example, Field et al. (2021) study the effect of an intervention aimed at increasing female's control of resources and find that its impact resulted in an increase in female labor supply, contradicting the implications of a standard collective model with individual utility depending on consumption and leisure. The paper's authors argue that, in reality, social norms play an important role in determining choices and this type of intervention might have led to a shift in such norms. The challenge then is to determine appropriate and validated measures of such norms.

Along similar lines, some interesting measures are those designed to capture what is sometimes defined as *social capital*, i.e. a set of norms that inform individual behavior and affect the ability of a society or community to provide public goods and internalize externalities, or other social attitudes that might affect individual interactions. A variety of measures, ranging from participation in certain activities (from blood donation to church attendance, see, for example, Guiso et al. (2004) and Guiso et al. (2006)) to data derived from field experiments (see Attanasio et al., 2012) to the effect of deterrence on preferences (see Cavatorta and Groom, 2020), have tried to capture social norms, attitudes, beliefs, and the role of culture.

In characterizing empirically certain markets (such as credit or insurance) and determining the model that best describes them, quantitative measures of specific frictions, jointly with *choice data*, can be very useful. In models of insurance with imperfect information, it may be useful to devise measures of the quality of information in a risk sharing group, as done, for instance, by Attanasio and Krutikova (2020).

What measures for what theories. We have discussed a few examples of theoretical models whose empirical analysis might need additional measures. One example is the identification of individual beliefs and preferences without strong assumptions.

In modeling parental behavior, for instance, it has been often assumed parents are fully aware of the nature of the process of child development. While such an approach can ease the analysis, a less restrictive model, embedding a more complex structure F in model (1), allowing for distortions in parental beliefs, may be more realistic and avoid misleading conclusions. In Section 4, to illustrate how additional and somewhat unconventional measures could (and should) be used in conjunction with traditional ones, we sketch a model of parental investment, which we analyze empirically in Section 6.

In some cases, the additional measures that allow the empirical characterization of more general models are just finer and more detailed versions of existing ones (as in the case of individual rather than the household level of consumption). In others, the new measures try to capture new concepts that are specific to the model being analyzed, such as (distorted) beliefs about child development or bargaining power within the household.

The lack of measures of key variables in the theoretical framework considered can force strong assumptions, and yield misleading results. These considerations are relevant, we believe, for the debate between Gul and Pesendorfer (2011) and Camerer (2011). While both papers make some interesting and important points, they take what we think is an over-restrictive approach. Gul and Pesendorfer (2011) insist that economic models describe the behavior and interactions of agents that are assumed to maximize a given objective function and that these models' features should be consistent with a set of axioms that help to frame them. The insistence on a framework consistent with theory is important, as it gives discipline and empirical bite to the theoretical models considered and, importantly, makes a well-defined welfare analysis possible. However, Gul and Pesendorfer want to refrain from using data and measures different from data on actual choice to characterize empirically these models. While the premise that the empirical models economists yaw should be theory consistent is a sound one, we believe that complex models are often better analyzed and characterized using data beyond those derived from observed choices. More importantly, such data could allow the use of richer models. Finally, while the measures of certain key factors might be affected by multiple biases, the problem, as clearly discussed by Manski (1990), is not with the measures *per se* but with the tools used to collect them.

Camerer (2011), on the other hand, points out that, again correctly in our opinion, new measures (such as the neurological data and biological markers he discusses) can be useful to better characterize individual behavior. While it is not obvious to us whether data from a *Functional Magnetic Resonance Imaging* (fMRI) could ever be related to specific aspects of individual behavior, such as discount factors or risk aversion, using measures different from choices, either bio-markers or others, to describe better individual behavior can be very useful, at the very least to improve the precision of our estimates and the efficiency of our tests. However, Camerer (2011) seems to want to characterize individual behavior in a way that abstracts from a set of theoretical axioms and to describe directly the relationship between biological mechanisms and behavior. Apart from the difficulty in pursuing such a strategy, such an approach goes beyond the realm of economics. Finally, the rejection of a specific model *in one context* is not a good reason to throw away the whole approach and work with models that are not consistent with a set of axioms.[17]

Interestingly, Benhabib and Bisin (2011) argue that traditional decision theory, as advocated by Gul and Pesendorfer (2011), focuses only on the need to model *choices*, while other approaches, somewhat misleadingly labeled as *behavioral economics*, want to understand the *processes* that lead to a specific set of choices. To better understand *processes*, additional measures, such as biological and anthropometric ones but also measures of a wide variety of latent factors, can be useful, if such measures are integrated in well-defined decision models of individual choices.

As economists, we need models that focus on economic ideas. Such models should not aim to describe completely psychological processes or define what happiness is. Well-constructed economic models, whose main aim is to describe and understand individual choices and interactions, should be based on a set of axioms and be consistent with them. To give empirical content to such models additional measures that complement data on choices (and prices and resources), whose design and features should be driven by the needs of the theory, can be useful.

3.3 Measurement systems

System (2) defines a *measurement system* in a fairly general way. Having discussed its relation to theory, we now move to a number of issues relevant for the specification and

[17] The discussion by Gul and Pesendorfer (2011) of present biases and the different approaches based on hyperbolic discounting (which leads to time inconsistent preferences) and a temptation model (which yields at the same time consistent preferences and present bias) is particularly illuminating in this respect.

estimation of such measurement systems. To make the discussion concrete, we use a specific characterization of the measurement system (2), similar to that used by Cunha et al. (2010). We denote with θ_{it}^{j} the $j-th$ element of the vector θ for individual i at time (or age) t. Let $m_{i,t}^{jk_j}$ be measure k_j of the K_j available and relevant for factor j. We assume that factors and continuous measures are related by the following system.

$$m_{i,t}^{jk_j} = \alpha_t^{jk_j} + \beta_t^{jk_j} \theta_{it}^{j} + \varepsilon_{it}^{jk_j}, \quad j = 1,...J; \;\; k_j = 1,...,K_j. \tag{3}$$

For discrete measures, we assume an Item Response Theory (IRT) model, extensively used in the psychometric literature and that we discuss at length in Appendix A1. For binary variables we have:

$$m_{it}^{jk_j} = \begin{cases} 1 & if \;\; \alpha_t^{jk_j} + \beta_t^{jk_j} \theta_{it}^{j} + \varepsilon_{it}^{jk_j} > 0 \\ 0 & otherwise \end{cases} \tag{4}$$

where $\varepsilon_{it}^{jk_j}$ in both equations are *additive* measurement errors independent from each other and from the latent factors. The binary case makes evident the role of the $\alpha_t^{jk_j}$ and $\beta_t^{jk_j}$ parameters representing the *discriminating* and *salience* properties of different measures. An item with very low or high values of α will have relatively low variability, and therefore will not be able to distinguish individuals with different values of θ, the factor of interest. Under some assumptions, the parameters of systems (3) and (4) (and of the distribution of the latent factors) can be estimated so to obtain estimates of the unobserved factors from the available measures.

Specification, identification and estimation of the measurement system. The relation between factors and measures as written in systems (3) and (4) is, in many ways, restrictive. However, it is useful to discuss some of the issues with measurement systems, starting with what is needed to identify empirically such a model.

The identification of the parameters of the model above or, more generally, that in system (2), requires the availability of several measures linked to each latent factor with measurement errors that are uncorrelated among at least two of them. An extensive discussion of the identification issues of these models is in Cunha et al. (2010), referring to the contributions of, e.g., Schennach (2004) and Hu and Schennach (2008).[18]

[18] In some cases, non-parametric identification requires at least 3 variables. *Identification* here refers to the parameters of the function g in system (2).

We will not repeat the formal arguments here, but we do stress that the need of independent measurement errors should inform the way surveys are designed and data collected. It would be desirable, for instance, to randomly assign different interviewers to collect different measures targeted at the same variables or collect them on different (randomly allocated) days. A similar argument applies to attrition (an extreme form of measurement error). One could, for instance, allocate resources spent on minimizing attrition randomly across observations. These considerations about data collection make clear the existing trade-offs: certain strategies might maximize data quality, while others might provide information that could be used to deal with measurement error.

Considering systems (3) and (4), it is clear that, even with specific assumptions about the distribution of the latent factors θ^j, not all parameters of these systems (that is, the $\alpha_t^{jk_j}$'s, $\beta_t^{jk_j}$'s, and the moments of the θ^j's), can be identified. It will be necessary to define a metric for the unobserved latent factors through appropriate normalization.

One possibility is to normalize the mean of the factors to 0 and the variance-covariance matrix to the unit one, an option often used in standard software packages, together with that of normality of the latent factors. Alternatively, one could normalize the $\alpha_t^{jk_j}$ and $\beta_t^{jk_j}$ of a specific measure to 0 and 1, respectively, effectively using that measure as the relevant metric for the unobserved factor. Both approaches are valid and effectively equivalent, with some important caveats.

Regardless of whether one normalizes the means and variance-covariance matrix of the factors or some of the parameters of the measurement system, when one has longitudinal data, it is necessary to establish whether one imposes these normalizations for the first t or for every t. Depending on the context, different approaches might be more useful. If one is interested in how the latent factors change over time,[19] it might be more useful to normalize only at one point in the observation sample. Such an approach, however, imposes the assumption that the relationship between the various measures and the latent factors is unchanged (i.e. measurement invariance). The imposition of these normalizations has implications for the interpretation of the evolution of different factors and their growth, as discussed in Agostinelli and Wiswall (2017).[20]

An issue, related to the normalization of the different measures and the identification

[19]For child development, for instance, one might be interested in measuring *children growth.*

[20]A related issue, relevant for longitudinal data, is when the measures that are appropriate for a factor change with time. These issues are discussed in the case of child development in Attanasio et al. (2020a).

of a metric for the unobserved latent factors, is that they often enter economic models as cardinal variables. In some cases, the relevant cardinal metric can be easily identified, in others that is not the case. This statement applies not only to models that consider, say, consumption or income but also, for instance, to the models of child development we consider below or in studies that consider the *value added* provided by schools. The issue is particularly difficult to deal with when the available measures are of an ordinal nature, using, for instance, Likert scales. The specification of the measurement system, in most contexts, should strive to obtain cardinal measures, which can be used in combination with ordinal measures but that provide the necessary anchor and metric that allows to obtain cardinal measures of the relevant factors.

As written in systems (3) and (4), the measurement system assumes that each factor affects only one measure; that is, it is a *dedicated* measurement system. This assumption can be relaxed and have several factors affecting a single measure. However, as mentioned by Cunha et al. (2010), to achieve identification it will be necessary to have at least one measure per factor that is affected only by that factor.

In many contexts, researchers estimate unobserved latent factors using a variety of tests that have become standard practice in the academic community and beyond. In the context of child development, for instance, much work exists in psychometrics which has influenced and shaped the development of a number of tests designed to measure different dimensions of child development or its drivers. Most of them are made of a large number of individual items that are routinely aggregated using scoring algorithms that deliver estimated developmental scores or indexes of parental investment. These algorithms were typically developed using factor models similar to those we discussed on samples of children on which the original items were tested. In other cases, the scoring mechanism is very simple, like the sum of correct answers to a number of questions.[21] While the original algorithms were eventually validated in a number of samples, it is not necessarily the case that, several decades later and in contexts that might be different from those in which the scoring algorithm was constructed, the same scoring algorithm is necessarily the best way to aggregate the information from the individual items. A feasible and more effective use of the data collected would be to

[21] An example in child development is the MacArthur Language Inventory test, a list of a few hundred words, with the child's caregiver informing whether the child understands or can say each of them. There is no reason to take the unweighted sum of such words as the estimated latent factor.

re-estimate the scoring algorithms (that is, the measurement system that relates latent factors to the available measures) in the new context where these data are collected. As we discuss in Section 5, this can also lead to the collection of different tests that could more easily be deployed in a given context. Along the same lines, new items could be piloted to complement the existing standard tests.

These issues are particularly relevant in developing countries, which are contexts very different from those where the tests were developed and the scoring algorithms designed, typically in developed countries. Many items might present flooring or ceiling effects, so that the specific tool is not able to capture any variability in the study sample. The estimation of a measurement system to construct context specific scoring algorithms is an effective and simple way to summarize efficiently the available measures.

Identifying causal links between latent factors. Another important set of related considerations is about the identification of causal links *between latent factors*. While the issue here is in terms of the theoretical structure that links the variables of interest, whether such links can be identified empirically might depend on the nature of the data that are available and collected. Indeed, the need to identify specific causal links should and often does drive the design and collection of specific surveys and data sets. A good example is the design of a Randomised Controlled Trial (RCT) where subjects are randomly exposed to a treatment or an intervention. In this particular example, the investigator creates exogenous variation (exposure to a treatment) on which information is collected to establish (if the experiment is performed correctly) a specific causal link. However, when the object of interest is more complex than the average impact of the intervention in a given context, the data collection strategy should be informed by the specific questions that researchers (or policymakers) might have. If one wants to extrapolate the results obtained in a given context to a different environment, or change details of an intervention, it will be useful and necessary to collect information on the drivers of individual behavior that might affect the outcome of interest.

In Attanasio et al. (2020b), for instance, the objective is to establish what mechanism generates the observed impact of a stimulation intervention on child development, with particular emphasis on the hypothesis that parental investment (in time and materials) played an important role. To find an answer to this question it is necessary to identify the causal link from parental investment to child development, in addition to the overall impact of the intervention on child development. One cannot use the treatment as a valid instrument to identify such a causal link, as the hypothesis of interest is whether the intervention has (or does not have) an impact on child development (directly or through other mechanisms). Attanasio et al. (2020b) use variation across different towns in prices of toys and other items as well as exposure to violence of mothers at the time when they were adolescents as determinants of parental investment that are assumed not to affect child development directly. This assumption (and the availability of the relevant measures), allow then to identify the relevant causal links and perform a *mediation analysis* that takes into account the fact that some of candidate mediators are variables determined by choice.

The general point we want to make is that the design of surveys should be informed by the specific needs and research questions that are being addressed. Survey instruments should use a variety of measurement tools and should not be restricted to collect information about individual choices. Identification of structural models could be achieved with much weaker assumptions by combining standard measures with, for instance, the elicitation of respondents' responses in hypothetical situations.[22] In our application in Section 6, we discuss how the empirical study of the model we present in Section 4 is made easier and more interesting by the construction of a different set of variables, some of which require the development of new measurement tools.

4 Modeling parental investment

In this section, we consider a specific application of the ideas we have discussed so far, sketching a theoretical model of household decision-making where some of its key components are often not directly observable. The main aim is to show how a number of variables that are not typically measured in standard surveys can be useful to identify the determinants of parental investment.

The framework we present is based on the collective model (Chiappori, 1988), which suggests that the household utility function is a weighted sum of the mothers

[22]The example of Juster and Shay (1964) is again a good one here.

utility function, $U^m(\cdot)$, and the father's utility function, $U^f(\cdot)$:

$$U = \mu(\mathbf{p}, Y; \mathbf{z}) U^m(\mathbf{c}, H(x)) + (1 - \mu(\mathbf{p}, Y; \mathbf{z})) U^f(\mathbf{c}, H(x)), \quad (5)$$

where **c** is a vector of consumption of both private and public goods, **p** is the corresponding vector of prices, and H is child human capital, dependent on investment in child development, x. $\mu(\cdot)$, often referred to as Pareto weights, represents the relative importance of the problem in equation (5) given to the mother's utility. The defining feature of this model is that, household choices, resulting from the interactions of different decision makers with possibly different preferences, are efficient, in that they maximize the function in (5), given certain constraints and a set of Pareto weights.

The model is silent about determining the Pareto weights. They are allowed to depend both on variables that enter the problem through the budget constraint, such as prices and income, and, crucially, other variables **z**, often labelled *distribution factors*. These are variables that matter for the sway in household decision-making but do not influence the budget constraint or utility functions directly (Browning et al., 2013).

The household faces two constraints: a standard (static) budget constraint and a *production function* of child development. In particular we assume that children human capital depends on initial conditions, H_0, parental investments, the financial investment in education, x, and other (unobserved) factors ε:

$$H = f(H_0, x, \varepsilon). \quad (6)$$

If the 'household' maximizes the utility function in equation (5), subject to the constraints we mention, parental investment will depend on household resources,on *individual* preferences as aggregated into *household* preferences by the Pareto weights $\mu(\cdot)$, and the properties of the assumed production function. We notice, however, that for the determination of parental investment, the properties of the production function of child development are not necessarily key, unless they coincide with those of the production function as *perceived by the parents*. In the presence of potentially distorted beliefs about the process of child development, what matters is the perceived productivity of parental investment, which can be different for husband and wives.[23]

[23]Note also that this framework can easily be extended to child development being a function of both material investment, x, and *time* investment, e.g., reading or talking to the child (see e.g., Attanasio et al.

A simple parametric example can make this framework clear and help us to relate it to the empirical exercises in Section 6. We assume that there are q private goods for husband and wife and that the only public good is child human capital. We also assume that both the individual utility functions and the production function of human capital are Cobb-Douglas (CD). While the commodities we consider are private, we allow the consumption of each spouse to affect the utility of the other spouse. Therefore, we have:

$$\ln U^i = \sum_{j=1}^{q} \left(\alpha^i_{jm} \ln C_{jm} + \alpha^i_{jf} \ln C_{jf}\right) + \alpha^i_k \ln H^k, \quad i = \{m, f\}; \tag{7}$$

where C_{jm} is the consumption of commodity j consumed by the wife, and C_{jf} is the consumption of commodity j consumed by the husband. α^i_{jm} and α^i_{jf} are the parameters of the CD function corresponding to the consumption of commodity j by each of the two household members in the utility function of member $i; i = m, f$. Analogously, α^i_k determines the utility derived by member i from the child's human capital. We assume that the utility function for the collective household is:

$$U = \mu U^m + (1-\mu)U^f. \tag{8}$$

The *perceived* production function for human capital is given by:

$$\ln H^k = \gamma^i_0 \ln H_0 + \gamma^i \ln X, \quad i = \{m, f\}; \tag{9}$$

where γ^i_0 and $\gamma^i, i = m, f$ are the parameters of the production function as perceived by the husband and wife. Finally, the budget constraint, with prices normalized to 1 for notational simplicity, is:

$$X + \sum_{j=1}^{q} (C_{jm} + C_{jf} + C_{jk}) = Y. \tag{10}$$

It is easy to show that the household parental investment function is given by:

$$X = \frac{\mu U^m \alpha^m_k \gamma^m + (1-\mu)U^f \alpha^f_k \gamma^f}{A} Y, \tag{11}$$

where the denominator is given by:

(2020b,c); Cunha and Heckman (2008); Heckman et al. (2013, 2020); and Todd and Wolpin (2003)).

$$A = \mu U^m \left(\sum_j (\alpha^m_{jm} + \alpha^m_{jf}) + \alpha^m_k \gamma^m \right) + (1-\mu) U^f \left(\sum_j (\alpha^f_{jm} + \alpha^f_{jf}) + \alpha^f_k \gamma^f \right). \quad (12)$$

While the expression for parental investment in equation (11) is particularly simple because of the CD assumption (which, for instance, implies unit elasticity and constant shares for all commodities), the expression illustrates the role played by the preference parameters (the $\alpha^i_{ji'},\ i,i' = m,f$), the Pareto weights, and the 'perceived' production function γ^i's. We note that mother's and father's individual preferences for each private commodity and child human capital are mediated by the Pareto weights given by μ.

The Engel curve of parental investment. While the CD assumption makes derivations very simple, the implied homotheticity is obviously not a plausible assumption. One possibility is to use a generalization of Deaton and Muellbauer (1980)'s AIDS system for Engel curves, as done, for instance, by Browning and Chiappori (1998). In this case, household i's budget shares for investment in human capital ($s_i = X_i/Y_i$) is a function of parental preference parameters, parental distribution factors, prices and total household expenditure, allowing the expenditure elasticity to be different from 1:

$$s_i = G(\tau_{\mathbf{i}}, \mu_i, \gamma_{\mathbf{i}}, \mathbf{p}) + \beta(lnY_i - a(\mathbf{p})) + u, \quad (13)$$

where G is a function that depends on the main determinants of the parental problem previously presented: parents' tastes for child human capital relative to other commodities, represented by two latent factors $\tau_{\mathbf{i}}$, parental beliefs about the productivity of parental investment in the production of human capital (again represented by two latent factors $\gamma_{\mathbf{i}}$), the relative bargaining power within the couple μ_i, the vector of relative prices, and, as in the AIDS system, log real income $lnY_i - a(\mathbf{p})$ and an unobserved shock u. It is also possible to make both the slope coefficient, β, and the price index, $a(\mathbf{p})$, functions of preference, τ_i, beliefs, $\gamma_{\mathbf{i}}$, and bargaining power, μ_i, and to extend this model to a Quadratic Almost Ideal Demand System (QUAIDS), as in Banks et al. (1997).

We will get back to these issues when, in Section 6, we estimate a simple version of equation (13) for investment in child human capital using measures on parental allocation preferences, parental decision-making power, and beliefs, in conjunction with actual choice data we collected in Tanzania. Some implications of this model, however, are clear. The share of expenditure on child development will be larger the higher the perceived returns to investment, and the more utility parents derive from child human capital over private consumption. If wives have a stronger preference for child development over private consumption than husbands, the investment share should be increasing in the weight wives have in decision-making. The opposite would hold true if husbands have a stronger preference for human capital than their wives.

5 New and old measures: using them jointly

In this section, we briefly describe a novel survey that was collected explicitly to pilot a number of new measures. In Section 6, we use (some of) these measures to get an empirical characterization of the model of parental investment we have sketched in Section 4. Here we provide a description of the different measures and how they were obtained. While the sample we use is relatively small, this data set is indicative of the possible use of a combination of new measures and choice-based data to estimate models of individual behavior. The description we provide in this section is also indicative of the difficulties that come with collecting new measures and the importance of a direct link between theory and the abstract constructs used in researchers' theoretical frameworks.

5.1 A pilot survey in Tanzania

The data we use in this and the next section were collected in two districts of the largely rural Kagera region in Tanzania. The overall goal of this data collection was to improve the measurement of child development and its drivers and design new measures. The survey included three samples: in the first, the respondents were mothers, in the second, fathers and, in the third, couples. The different samples are useful to characterize differences across mothers and fathers and how joint (couple) decisions are made. We describe the data and their collection in detail in Appendix A2. Here we focus on the novel measures. Unfortunately, the sampling strategy made the three samples somewhat unbalanced, in particular with respect to age. However we still find it useful to compare them here for illustration purposes.

Bargaining Power. The power that women have within households has received much attention, especially when analyzing models of intrahousehold allocations. In our Tanzania sample, a first measure related to the bargaining power within the couple replicated the approach used by Almås et al. (2018), who conducted a controlled ("laboratory") experiment in a sample collected within an RCT in North Macedonia. The measure was designed to capture the potential impact of targeting women rather than men with a Conditional Cash Transfer, which was given to women in some villages and to men in others. To capture the *bargaining power* latent factor, after the initial data collection, the wives were called to an office to run an incentivized experiment. They were told: "Here are 100 Denars that we will give to your husband. How much are you willing to pay to have them paid to you?" This amount, which is completely independent of the government-administered cash transfer, was actually paid out (to the husband or the wife, depending on her choice).[24] A second hypothetical question was asked considering much larger amounts. The idea is that women willing to sacrifice a higher proportion of the amount offered are less powerful within the couple.

Almås et al. (2018) show that such a measure of bargaining power, which we label *Willingness To Pay* (WTP), correlates with a number of observables in a predictable manner. Moreover, in villages where the government grant was targeted at wives, the WTP decreased significantly; in these villages, women were willing to pay less to get control of any additional transfers. The incentivized and the hypothetical exercises in Macedonia yield similar results, indicating that hypothetical formulations of these questions could also be used in surveys.

In the Tanzanian sample, a hypothetical version of the WTP elicitation was conducted. Furthermore, unlike in North Macedonia where the question was asked only to wives, in Tanzania it was also asked to husbands in the *fathers sample*. In Table 1, we report the average share of the 6600 TSH that the respondents (i.e., the wives in the mothers and couples samples and the husbands in the fathers sample) were willing to forfeit so that the payment would be made to them rather than their spouse. As in Almås et al. (2018), we interpret this share as being inversely related to the control of resources the respondent has on the household resources, which in turn can be related to the Pareto weights of the collective model discussed in Section 4.

[24] 100 Denars corresponds, for this sample, to two days of paid work.

Table 1: Willingness to pay (out of 6600 TSH)

	Wives	Husbands
	Mothers and Couples samples	Fathers sample
Average share	0.320	0.100
Median share	0.061	0.008
Std. Dev.	0.394	0.258
Observations	215	98

Note: The table displays the descriptive statistics for the Willingness To Pay (WTP) variable. The WTP is measured as the share of the 6600 TSH that the respondents (i.e., the wives in the mothers and couples samples and the husbands in the fathers sample) were willing to forfeit so that the payment would be made to them rather than their spouse. In the mothers sample, the statistics are computed over the households where the father is present. Source: Tz Pilot.

The distribution of WTP is very skewed, with a few observations with very high values. This feature of the distribution is reflected in large differences between median and mean of the distribution. We observe a considerable difference between husbands' and wives' WTP, with wives willing to sacrifice on average 32% of the transfer to get control over it. The median wife, both in the mothers and couples samples, is willing to pay just over 6% of the transfer. Husbands, on the other hand, are willing to pay only 10% on average, with the median being 0. As we mention above, the three samples we are considering (in particular the fathers sample) are not strictly comparable, because of the age differences induced by the sampling design. However, the difference in the WTP is very marked and probably reflects different bargaining positions within the marriage, indicating that men have more control over resources than women.

Translating the WTP measure into a measure of bargaining power within the couple is not simple. To perform such an exercise, it would be necessary to identify a mapping from the theoretical constructs of the collective model, such as the Pareto weights, and individual preferences with the possibility of altruism and public goods or some of its intermediate outcomes, such as *sharing rules*, to the WTP measure.

Existing surveys provide alternative information on resource control within the household through a number of questions that explicitly ask who in the households is responsible for decisions on expenditure on various commodities and other decisions, with the possible answers being 'the husband' 'the wife' or 'both'. In the Tanzania survey we are using, a number of questions, taken from the Tanzania Demographic and Health Survey (DHS), were posed to the respondents. The questions were about major household expenditures, children's education, health expenditures, what food to cook, and whether the wife can go out.

In particular, the respondents were asked who is mainly responsible for a number of decisions, including major household expenditures, children's education, health expenditures, what food to cook, and whether the wife can go out. To avoid the often-observed bunching around the 'both' answer, when this option was chosen the respondent was asked who had the final say.

Table 2: Measurement system for decision making

	Decision making			
	Wives		Husbands	
	Mothers and Couples sample		Fathers sample	
	β	α	β	α
Own health	1.000	0.000	1.000	0.000
	-	-	-	-
Children's health	1.283	-0.171	1.213	-1.140
	(0.272)	*(0.170)*	*(0.484)*	*(0.249)*
Children's schooling	1.167	-1.261'	1.021	-0.680
	(0.381)	*(0.181)*	*(0.419)*	*(0.194)*
Household purchases	0.662	-0.988	1.506	-0.386
	(0.219)	*(0.132)*	*(0.638)*	*(0.242)*
Cooking	0.278	0.904	0.563	-1.039
	(0.120)	*(0.117)*	*(0.296)*	*(0.167)*
Visiting	0.639	-0.932	0.582	0.235
	(0.198)	*(0.129)*	*(0.302)*	*(0.138)*
Factor mean	0.468		-0.217	
Factor variance	0.882		0.402	

Note: This table shows the loading factors (β) and the intercept (α) of an IRT estimated from questions about who is mainly responsible for a number of decisions within the household, including: major household expenditures, children's education, health expenditures, what food to cook, and whether the wife can go out. Columns (1) and (2) are for the mothers and couples sample, and Columns (3) and (4) for the fathers sample. Standard errors in parentheses. Source: Tz Pilot.

Using the answers to these questions in the mothers and couples samples (where the wife is the respondent) and the fathers sample (where the husband is the respondent) we estimate a measurement system and, from that, the latent factor of interest, reflecting women's decision power in the couple. The estimates of the measurement system parameters are reported in Table 2.

We can now relate the WTP measure to the 'decision making' factor extracted from more traditional measures of control of resources. In particular, in Table 3, we report the results of a regression of the WTP on the decision making latent factor estimated first in the sample of mothers and couples and then on the sample of fathers.

We limit the sample of mothers to the households where the husband is present and allow, in the first regression, an intercept shift for the couples sample. We find that in the mothers and couples samples, the two variables are significantly and negatively related (the higher the share respondents are willing to forfeit to get control of the payment, the less the mothers' decision making power within the household), and the couples sample shift is not significant. The R-squared of this regression, however, is very low, indicating that there is a considerable amount of variation in WTP which is not related to the traditional measures of control. A similar result applies to the sample of fathers, although the coefficient is not significant.

Table 3: Willingness to pay and decision making factors

	Willingness to pay	
	Wives	Husbands
	Mothers and Couples samples	Fathers sample
Decision making factor	-0.079**	0.071
	(0.033)	*(0.054)*
Couple	-0.041	
	(0.054)	
R-squared	0.030	0.018
Observations	215	98

Note: The table displays regressions of the willingness to pay on the decision making variable and a dummy for the couple treatment arm of the experiment. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: Tz Pilot.

Jayachandran et al. (2021) have used an approach similar to the one we have used to derive the WTP described above in India. They also use a machine learning algorithm to identify questions that accurately reflect women's decision power in the couple. They conclude that, in the Indian context, the latter approach seems to work better.

Beliefs on returns to parental investment. Another important driver of individual choices that we consider in our application is parental perception about the process of child development. As is clear from the model presented in Section 4, parental investment is driven by parental perception of the return on child development. While much of the existing literature assumes that parents know the process of child development and how it depends on the child's current development, parental investment, and possibly other factors, it is increasingly clear that these perceptions might be distorted.

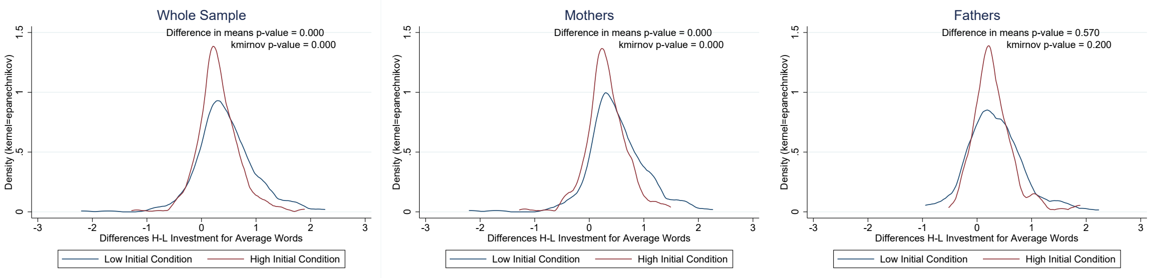
Several studies, which we mention in Section 3.2, have started eliciting beliefs about the relationship between parental behavior and child development. We use an approach similar to Attanasio et al. (2019b), who measure mothers' beliefs about the process of child development within a survey of an RCT evaluating a parenting intervention.

The approach consists of presenting mothers with *scenarios* in terms of initial conditions and investment and asking them to map these scenarios into child development outcomes. The implicit assumption is that mothers use the same mapping between latent factors and observable markers, so the scenarios proposed in the questionnaires have a relation to the latent factors that researchers want to capture. This approach allows researchers to estimate perceived rates of return to parental investments under different initial conditions.

In the Tanzania sample, we use this approach to elicit beliefs – for both fathers and mothers – about different aspects of the developmental process and the importance of certain parental inputs. We use the answers to the beliefs questions to estimate a measurement system and extract a 'beliefs factor', which we can use to estimate the perceived return to parental investment. The estimates of the measurement systems are reported in Table A2 in Appendix A3. The returns we consider for high and low levels of initial development are measured as the difference in the expected outcomes between high and low levels of investment for the two levels of initial conditions.

In Figure 1, we plot the distribution of expected returns under low and high initial conditions for the whole sample as well as for the mothers and fathers sub-samples. Figure 1 also reports a test for the difference of the distribution means and a Kolmogorov-Smirnov (KS) test for the difference between the two distributions. For beliefs about language development, returns to parental investment are perceived to be higher for low than for high initial conditions in the whole sample and for mothers (in the mothers and couples samples): the difference in means for the whole sample is equal to 0.140 (p-value=0.000) and to 0.200 (p-value=0.000) for mothers'. For the fathers sample there is no significant difference between the returns with high or low initial conditions: the point estimate of the difference is 0.030 (p-value=0.570).

Figure 1: Beliefs on language development: Returns to parental investment

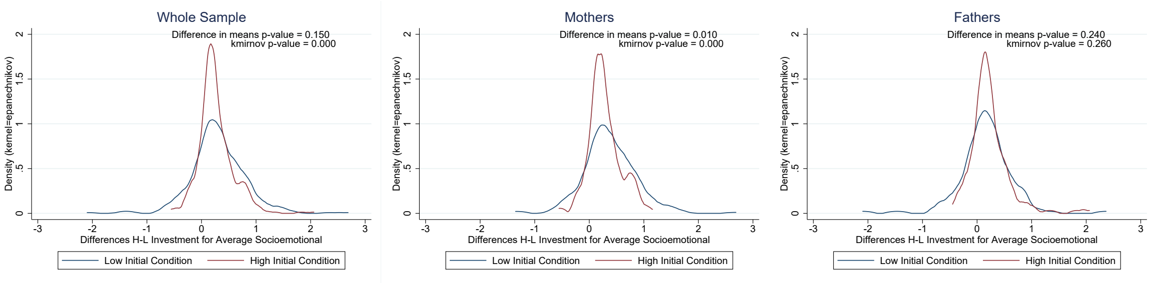


The first two facts are qualitatively consistent with the evidence from Colombia reported in Attanasio et al. (2019b): poor parents seem to think that parental investment is more productive and effective at low levels of initial development. More generally, the entire distribution seems to be different, with the low initial condition returns presenting more dispersion and shifted to the right. Mothers have a higher expected return to investment for low initial condition children than fathers, as the difference is equal to 0.175 (p-value=0.000). There is no significant difference between fathers and mothers on expected returns for high initial condition children.

In addition to the beliefs about cognitive development, we perform a similar exercise to measure beliefs about the effect of parenting on socio-emotional development, a type of belief that has not been measured before. In Figure 2, we report the distributions of expected returns from different initial values in the whole sample and, as for beliefs about language development, for the mothers and fathers samples separately. As in Figure 1, we report KS tests for the difference between the two distributions and a test for the difference between the low and high initial conditions distributions.

The results are similar to those on the beliefs about the effect of parenting on cognitive development and language. In the whole sample, returns to parental investment on socio-emotional development are perceived to be higher for low than for high initial conditions: the difference in means for the whole sample is equal to 0.044 (p-

Figure 2: Beliefs on socio-emotional: Returns to parental investment



value=0.150). This effect is coming from mothers, where the difference is equal to 0.097 (p-value=0.010). We see no such effect in the fathers sample, where the difference in means is equal to -0.063 (p-value=0.240).

Measuring preferences. As discussed above, an important determinant of individual choices is preferences. While, under a number of assumptions, the structural parameters that characterize them can be estimated from choice data, in some contexts, answers to hypothetical questions can be particularly useful. In a recent paper, Ameriks et al. (2020) use stated hypothetical choices elicited with Strategic Survey Questions (SSQs) to estimate some preference parameters in a fully specified structural model. The authors note that to develop direct measures of preferences, it is desirable to develop survey instruments that allow respondents to provide information that identifies preference parameters in a language they are comfortable with as well as in a format that allows for a precise mapping to the structural parameters of interest.

In studying the model of parental investment in Section 4, in which preferences of different decision makers interact to determine household choices, it is important to elicit not only preferences on the allocations of scarce resources among different commodities but also between different individuals within the family. Cherchye et al. (2021) use a similar approach in Kenya to elicit tastes for couples' allocation between child investment and own consumption.

In the Tanzania sample, before the standard survey, a key respondent or respondents (the wife, the husband, or the couple jointly) were asked to allocate a hypothetical amount, represented by a pile of beans, between different expenditure categories and between three individuals (the husband, the wife, or the child in the household).[25] The six possible expenditure categories considered were: clothing, food, learning materials (such as books, notebooks, and pens), health expenditures, transportation, and school expenditures. These categories were chosen to be able to match the information collected on actual expenditure. While the question was not explicit about this issue, we interpret the answers to the hypothetical questions as referring to the allocation of additional resources that the household would normally have access to.

Table 4: Average share of expenditure allocated to household members

	Mother decision (s.e.)	Father decision (s.e.)	diff (p-value)	Couple decision (s.e.)	diff (p-value)
To self	0.268	0.257	-0.01	0.250	-0.02
	(0.008)	*(0.007)*	*(0.348)*	*(0.009)*	*(0.148)*
To spouse	0.175	0.219	0.04	0.232	0.06
	(0.009)	*(0.007)*	*(0.000)*	*(0.009)*	*(0.000)*
To child	0.558	0.524	-0.03	0.518	-0.04
	(0.011)	*(0.013)*	*(0.043)*	*(0.013)*	*(0.019)*

Note: This table shows the average share of expenditure to household members for the different samples. Standard errors are in parentheses. The p-values refer to the test of the difference between the mothers and fathers samples and the mothers and couples samples. Source: Tz Pilot.

In Table 4, we report the share of the total additional resources allocated to each individual in the family (spouse, child, and self) for each of the three samples. A number of interesting results emerge from this exercise. First, despite the good considered being private, both mothers and fathers allocate some resources to their spouse, indicating that the participants care about their spouse's consumption. Second, mothers allocate more than fathers to children, but allocate the same share to themselves, which implies mothers allocate less to their spouses than fathers. Third, the couple's decisions seem much closer to those of fathers than those of mothers. The similarity between the fathers and the couple allocations might indicate large differences in decision-making power

[25] The question was posed as: "We would now like to understand how you would prefer to spend 300k Shillings, if we were to give this money to you. Use these 60 beans, each representing 5k Shillings, and this cardboard card with 3 different expenditure options (mother, father, and your child); for each question distribute the beans according to your preferences. Imagine that your child is 5 for this exercise." See Almås et al. (2020a) for a full description of the protocol followed.

between spouses as fathers, consistent with the evidence on the WTP, hold considerably more decision power within the household.

Next, we look at the allocation among different commodities, and in particular, for the resources allocated to the child. In Table 5, we report the shares allocated to the child split into the six different commodities. There are some differences between mothers and the other samples, particularly in clothing and health (where the mothers' shares are significantly higher) and in learning materials (where the fathers' share is marginally higher). Once again, the couple's decisions are more similar to those of fathers than those of mothers.

Table 5: Allocation to the child

	Mother decision (s.e.)	Father decision (s.e.)	diff (p-value)	Couple decision (s.e.)	diff (p-value)
Clothing	6.628	5.559	-1.07	5.493	-1.13
	(0.225)	*(0.311)*	*(0.005)*	*(0.195)*	*(0.000)*
Food	6.062	5.338	-0.72	4.401	-1.66
	(0.302)	*(0.269)*	*(0.076)*	*(0.251)*	*(0.000)*
School exp.	7.434	7.529	0.09	7.282	-0.15
	(0.353)	*(0.573)*	*(0.886)*	*(0.456)*	*(0.791)*
Learning mat.	5.503	5.213	-0.29	5.697	0.19
	(0.247)	*(0.285)*	*(0.441)*	*(0.317)*	*(0.629)*
Health exp.	5.159	5.213	0.05	5.761	0.60
	(0.207)	*(0.252)*	*(0.866)*	*(0.232)*	*(0.054)*
Transportation	2.683	2.603	-0.08	2.430	-0.25
	(0.182)	*(0.202)*	*(0.769)*	*(0.199)*	*(0.349)*

Notes: This table shows the descriptive statistics of allocation of expenditure on children. Standard errors are in parentheses. The p-values refer to the test of the difference between the mothers and fathers samples and the mothers and couples samples. Source: Tz Pilot.

In the model we discussed in Section 4, it was clear that the relative parental taste for child human capital and alternative allocations of resources is a key determinant of parental investment. While we do not estimate a structural version of that model, as it would be implied, for instance, by the AIDS version of Engel curves in equation (13), we use the answers about stated preferences in the couple samples to derive some information about individual couples' tastes and relate them to parental investment. As is evident from equation (11), which was derived under homothetic preferences, parental investment can depend on the shares of resources allocated to adult goods relative to the resources allocated to child consumption and parental investment.

To estimate a factor representing *the taste for child human capital*, we use the answers to the allocation questions to construct eight variables as the ratios of the resources allocated to each spouse for four adult commodities (food, clothing, health, and transportation) to the resources allocated to the child and estimate a factor model to extract a latent factor we label *relative taste for child human capital* from these eight variables. We perform this analysis in each of the samples and report the loading factors for each of the variables, as well as the intercept for the equations of the measurement system corresponding to each observable variable, in Table A3 in Appendix A3. The factor estimated from this analysis is what we use in Section 6 to model parental investment. From Table A3, several variables seem to be important markers of the taste for child development. Furthermore, we find important differences in the three samples, with the preferences in the couples sample seeming more similar to the fathers' rather than the mothers'. In the latter, measures of the mother's and father's health expenditures and father's food expenditures are not particularly important.

Having presented some evidence on the new measures that were collected in Tanzania, one important issue and challenge is their validation. In particular, we check if these measures co-vary in a sensible way with choice data or, more generally, with standard measures. This would be a first step towards a systematic use of these measures within models of individual behavior. We turn to this in the next section.

5.2 Using available measures efficiently

Even when the latent factors that one wants to analyze are reasonably standard, it may not be easy to obtain meaningful measures from the available raw data that can be used in empirical analysis. In many contexts, certain algorithms that aggregate available measures become the accepted standard and are widely used, even when alternative approaches may be more efficient and meaningful. As discussed in Section 3.3, the explicit specification of a measurement system can be interpreted as the construction of a scoring mechanism that efficiently uses the available data to estimate the latent factors of interest. The study of child development and its drivers, such as parental investment or school quality, provides a particularly salient example that has received substantial attention in recent years.

It is recognized that measuring child development is difficult, especially in the early years and when one wants to assess different dimensions of development, including socioemotional skills. Analogous considerations apply to measures of the drivers of child development, such as parental investment and school quality. These difficulties are even more serious in developing country contexts, both because the administration of some of the most frequently used tests is often difficult and requires specialized testers and because many of these tests were developed and validated in what has been termed Western, Educated, Industrialised, Rich, and Democratic (WEIRD) samples (Henrich et al., 2010).[26] While a number of sophisticated tests have been developed and validated in developed countries, the relevance and effectiveness of such tests in completely different contexts might be limited.

Because of these considerations, a number of efforts to develop a new generation of child development tests are under development.[27] Within the Tanzania project we have been discussing, a new test was also developed, described in Attanasio et al. (2022), which combined elements from different well-established tests (such as the Bayley Scales of Infant and Toddler Development, Third Edition - Bayley-III, the Caregiver Reported Early Development Instruments (CREDI), and others) to construct an efficient and easy to administer test with a limited number of items. The approach consisted of estimating a measurement system such as in systems (2) and (3), relating the dimensions of interest to the various elements that make up these tests. This procedure, now widely used (see, for instance, Cunha et al. (2010), Agostinelli and Wiswall (2017), Heckman et al. (2020) among others) can be seen as an effective alternative to the use of standard algorithms that typically come with these tests. The construction of a new scoring algorithm through the estimation of a measurement system obtained in a given context is an effective way to adapt the existing tests to new realities and make different

[26]For concepts such as parental investment and school quality, the application of tests developed in different contexts can easily result in flooring and ceiling effects.

[27]The UN has facilitated the development of the Multiple Indicator Cluster Survey (MICS), which has recently been discussed in Bornstein et al. (2021). The Gates Foundation has funded a large effort to pull together a number of new indicators that could be comparable across contexts and countries and would be relatively easy to administer. This has given rise to the Global Scale for Early Development (GSED) initiative, discussed in GSED (2021) and Black et al. (2019). McCoy et al. (2021) describe the construction and use of the CREDI questionnaire.

contexts comparable.

Attanasio et al. (2022) use the estimates of the measurement system to identify the most informative items to measure different dimensions of child development. Attanasio et al. (2022) show that the use of the most informative items yields estimates of child development that contain virtually the same information as the complete tests and are much cheaper and quicker to collect, especially in developing countries.[28]

Table 6: A measurement system for parental investment

	Couples sample	
	β	α
Parental Activities	1.000	0.000
	-	-
Play Material	0.362	-0.201
	(0.136)	*(0.086)*
Material Investment	1.007	0.201
	(0.377)	*(0.176)*
Expenditure on	0.351	-0.080
children share	*(0.128)*	*(0.070)*
Social scale	0.657	0.181
	(0.329)	*(0.173)*
Didactic scale	0.415	0.031
	(0.235)	*(0.136)*
Factor mean (variance)	-0.376	(0.436)

Note: This table shows the loading factors (β) and the intercept (α) for markers of parental investment. Columns (1) and (2) are for the couples sample. Standard errors in parentheses. Source: Tz Pilot.

A similar set of issues is relevant for measures of latent factors that are key drivers of child development, such as parental investment. It is not always clear how many dimensions of investment to consider and how to adapt available measures to different contexts.[29] A number of standardized measures exist and have been widely used,

[28]As we do not use these data in the exercise in Section 6, we do not report here the details of these results and refer the reader to Attanasio et al. (2022).

[29]Researchers use different strategies to measure parental investment. Attanasio et al. (2020b), for instance, consider time and material investment separately and show that they might have different impacts on child development. Others, such as Cunha et al. (2010), consider a single dimension.

such as the Home Observation Measurement of the Environment (HOME) index or the Family Care Indicators (FCI). However, given the set of items that make up these tests, it is not clear that the same scoring algorithm should be used in different contexts, as different items might be differently salient and relevant depending on the context.

In the Tanzania context considered here, we follow an approach to measuring child development similar to that taken by Attanasio et al. (2022) and estimate a factor model that identifies a single latent factor. We use this factor as our measure of parental investment in Section 6. In particular, we use a number of items as markers of parental investment, including: (i) time spent in activities with children; (ii) play materials present in the house; (iii) material investments for the child (including food, clothing, footwear, confectioneries, among others); (iv) share of expenditure on children items over the total expenditures of the household; (v) items from the *social scale* from the Parental Style Questionnaire (PSQ; Bornstein (1996)); and (vi) items from the *didactic scale* from the PSQ. In Appendix A3.3 we define (i)-(vi) and Table A4 reports descriptive statistics on each components of parental investment.

In Table 6, we report some of the estimates of the parameters of systems (3) and (4) for the sample of couples. We notice that several markers are relevant to the factor we are considering. Different measures of parenting skills, for instance, do not seem particularly relevant for the parenting investment factor, while material investment plays an important role.

6 A model of parental investment

In the model we presented in Section 4, the Engel curve for parental investment can be seen as an approximation similar to the one used by Browning and Chiappori (1998). In this section, we estimate a version of equation (13) for parental investment using both choice data on investment and total expenditure, as well as the measures of preferences, beliefs, and bargaining power that we collected and discussed in Section 5. Rather than estimating a detailed version of the model that incorporates tastes for a variety of different commodities, we focus on parental investment and summarize the information on couples' preferences for child development relative to other uses of resources. We want to investigate how our novel measures co-vary with actual behavior in terms of parental investment in a way which is consistent with the model we presented.

We presume that in the sample where the non-conventional questions are directed to the couple, the answers reflect the 'aggregated' couple preferences and beliefs about child development. Therefore, for this exercise, we use only the couple samples, where questions are directed to the couple, as it would be hard to model observed investment by the couple on the basis of the preferences and beliefs of only one partner.

We estimate different measurement systems to extract from the available measures information about factors that enter the model in equation (13). In Section 5, we discuss the systems we use to extract the latent factors representing couples' relative tastes for child development, their beliefs about the productivity of parental investment, and bargaining power within the couple as well as actual parental investment. The estimates of these measurement systems are reported in Appendix A2.

In equation (13), the share of parental investment in total expenditure depends on (the log of) total expenditure and the latent factors representing both spouses' tastes and beliefs as well as a bargaining power factor, which aggregates the individual factors. As we interpret the taste and beliefs factors elicited in the couples sample as reflecting the *couples' preferences and beliefs*, one can modify equation (13) as:

$$s_i = G(\tau_i(\mu_i), \gamma_i(\mu_i), \mathbf{p}) + \beta(lnY_i - a(\mathbf{p})) + u_i \tag{14}$$

where τ_i and γ_i are now unidimensional factors that aggregate (through the bargaining power factor μ_i) the tastes and beliefs of the two spouses. If we interpret the responses from the couples sample as reflecting these aggregate factors, the bargaining power should not enter the parental investment equation once we control for the aggregated factors. However, it is possible that the linear specification that approximates equation (13) is too restrictive so that the bargaining power factor could enter it significantly.

In Table 7, we report the results of regressing the parental investment factor on: log total expenditure, the beliefs and taste factors (and their interactions), and our measure of bargaining power. As parental investment is a factor with an arbitrary scale, the size of the coefficients in Table 7 is difficult to interpret. However, we notice that (log) total expenditure attracts a positive and significant coefficient in all specifications.

In column 1, we only use choice data, in that we regress the investment factor (which is determined by investment in time spent with children and expenditure in commodities targeted to children) on the log total expenditure only. In column 2, we introduce our estimates of taste and beliefs questions to capture the determinants of parental investment as function $G()$ in equation (14). We observe that the relative taste for other commodities is strongly significant and with the expected negative sign. This result can be partly interpreted as a validation of the measure of taste that we use, which is derived from an experiment on hypothetical allocations; the specific experiment does not refer to actual investment choices at all. Analogously, the factor measuring beliefs about the productivity of investment is also strongly significant with the expected positive sign in the equation for parental investment.

Table 7: Modeling expenditure: Couples sample

	(1)	(2)	(3)	(4)	(5)
	Exp share	Exp share	Exp share	Exp share	Exp share
Log of total expenditure	-0.097***	-0.082***	-0.080***	-0.080***	-0.079***
	(0.030)	*(0.027)*	*(0.027)*	*(0.027)*	*(0.027)*
Bargaining Power (BP)		-0.008	-0.009	-0.009	0.054
		(0.062)	*(0.063)*	*(0.064)*	*(0.105)*
Relative taste for child human capital (RC)			-0.137	-0.143	0.146
			(0.179)	*(0.331)*	*(0.421)*
Beliefs			0.035	0.032	0.040
			(0.121)	*(0.198)*	*(0.122)*
Beliefs*RC				0.023	
				(1.118)	
BP*RC					-0.362
					(0.488)
R-squared	0.142	0.202	0.208	0.208	0.211
Observations	142	126	126	126	126

Note: The table displays regressions using the factor of expenditures share as a left hand side variable and the following right hand side variables: i) Log of total expenditure is the logarithm of total household expenditure from the household survey; ii) Bargaining power is a measure of female bargaining power, and more precisely, it is 1 minus the share that the woman is willing to pay to gain control over a fixed amount of money from our experiment; iii) Relative taste for child human capital is a factor of ratios of the resources allocated to each spouse for four adult commodities to the resources allocated to the child in our allocation experiment; and iv) Beliefs is a factor of the returns to parental investment on Language and Socio-emotional skills elicited in our survey. All models are controlled by child and mothers characteristics: child's age and gender, number of siblings in the household, dummy for mother's secondary education and mother's cognition skills measured by the Raven's test. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: Tz Pilot.

In column 3, we add to the variables in column 2 the bargaining power factor. As mentioned above, if we interpret the answers given by the couple's questions on their tastes and beliefs as representing the aggregation of the individual tastes and beliefs, (and if the equation is linear), such a variable should not be a significant determinant of investment. We find that, while the other coefficients do not change much with the introduction of this variable, representing the wife's power within the couple, takes a negative coefficient which is significantly different from zero. Without a further detailed analysis of the various factors at play, it is difficult to interpret this coefficient.

As the significance of the bargaining power factor might be signal the presence of nonlinearities in the function $G()$ in equation (14), in columns 4 and 5 of Table 7, we introduce interactions of the taste factor with beliefs (in column 4) and with bargaining power (in column 5). In neither of the two columns do we find any significant interactions between taste and the two variables considered.

This exercise is a first step in utilizing novel measures in a unifying framework that combines elicited beliefs, preferences, and decision-making power with observational data. This research's next step is to introduce these measures into a structural model of parental behavior and map them more directly onto its parameters. To achieve such a goal, it may be necessary to develop finer measurement tools than those used here.

7 Conclusions

In this paper, we have analyzed the role that measurement does and should play in economics and its relation to economic theory. We argued that measurement issues – including what should be measured, how to construct effective measures of the latent factors that populate economic models, and how to use such measures – should be informed by economic theory. Economic models are attempts to describe certain aspects of human behavior in specific contexts in a coherent fashion that allows generalizations, extrapolation, and ultimately the identification of causal links between different variables. Depending on what is being modeled or 'explained,' bringing the relevant latent factors to data might require the measurement of different variables.

Academic economists have, for a long time, shied away from measuring certain variables. With some exceptions, they have relied almost exclusively on data on choices, prices, and resources, or, more generally, objectively observable variables. They refrained from using measures of attitudes, intentions, stated preferences, beliefs, subjective expectations, social norms, etc. There are good reasons to treat such measures with caution and even skepticism as they might be difficult to collect and can be affected by different types of bias. However, empirical work that relies exclusively on choice data, and supposedly objective measures, imposes strong restrictions to the economic theories and models that can be brought to data, which typically take the form of strong assumptions on the structure of the models one works with. Many important models, some of which we discussed, could be analyzed with much more substance and empirical bite were they to use a wider set of measures. Such measures, while not widely used by economists, have been extensively used in other disciplines, from marketing to psychology and child development.

Obviously, given the difficulties in collecting the innovative measures we are advocating, they should be validated properly. Moreover, as we have argued, these measures are particularly useful when utilized in combination with choice data and other standard measures. This combination could be used both to validate the new measures, possibly via simple correlations and to estimate and test richer models.

In the second part of the paper, we have put in practice some of these ideas, using a set of new measures collected in rural Tanzania to estimate a model of parental investment, which we relate to measures of parental preferences, elicited from stated preferences, bargaining power within couples, and parental beliefs about the process of child development. While we do not estimate a fully structural model, we show how these data can be used in combination with standard measures to quantify the importance of different factors affecting parental behavior. In a next step, these data can be used to identify important causal links. Eliciting from respondents' information about choice under different counterfactual and hypothetical scenarios gathering information about preferences, one can, in some contexts, solve *through measurement* some of the endogeneity and identification issues that make empirical work challenging.

Every day, new data are being created and used, most noticeably administrative data from a variety of sources and contexts. This is obviously a positive development. However, we believe that well-designed and innovative survey measures of variables and constructs that are theoretically relevant can be just as useful. Indeed, many new measures are being developed in this direction, including those we have cited. Future research should devote substantial efforts to develop, design, implement, and validate new measurement tools that can provide useful evidence for economic theory and, ultimately, public policy.

References

Acemoglu, D. and D. H. Autor (2011). "Skills, tasks and technologies: Implications for employment and earnings." In "Handbook of Labor Economics Volume 4," (edited by Ashenfelter, O. and D. E. Card). Amsterdam: Elsevier.

Acemoglu, D. and P. Restrepo (2019). "Automation and new tasks: How technology displaces and reinstates labor." *Journal of Economic Perspectives* **33**(2), 3–30.

Ackerberg, D. A., K. Caves, and G. Frazer (2015). "Identification properties of recent production function estimators." *Econometrica* **83**(6), 2411–2451.

Adams, A. and A. Andrew (2019). "Preferences and beliefs in the marriage market for young brides." Tech. rep., IFS Working Papers.

Aghion, P., A. Bergeaud, T. Boppart, P. J. Klenow, and H. Li (2019). "Missing growth from creative destruction." *American Economic Review* **109**(8), 2795–2822.

Agostinelli, F. and M. Wiswall (2017). "Identification of dynamic latent factor models: The implications of re-normalization in a model of child development." Tech. rep., NBER, WP No 22441.

Alesina, A. and G.-M. Angeletos (2005). "Fairness and redistribution." *American economic review* **95**(4), 960–980.

Alesina, A. and E. La Ferrara (2005). "Preferences for redistribution in the land of opportunities." *Journal of Public Economics* **89**(5), 897–931.

Alesina, A., A. Miano, and S. Stantcheva (2018a). "Immigration and redistribution." Tech. rep., NBER, WP 24733.

Alesina, A., S. Stantcheva, and E. Teso (2018b). "Intergenerational mobility and preferences for redistribution." *American Economic Review* **108**(2), 521–54.

Almås, I. (2012). "International income inequality: Measuring ppp bias by estimating engel curves for food." *American Economic Review* **102**(2), 1093–1117.

Almås, I., A. Armand, O. Attanasio, and P. Carneiro (2018). "Measuring and changing control: Women's empowerment and targeted transfers." *The Economic Journal* **128**(612), F609–F639.

Almås, I., O. Attanasio, P. Jervis, and C. Ringdal (2020a). "Targeted cash transfers and children's welfare: Should women be targeted?" Tech. rep., NHH working paper.

Almås, I., A. W. Cappelen, and B. Tungodden (2020b). "Cutthroat capitalism versus cuddly socialism: Are americans more meritocratic and efficiency-seeking than scandinavians?" *Journal of Political Economy* **128**(5), 1753–1788.

Almås, I. and Å. A. Johnsen (2018). "The cost of a growth miracle–reassessing price and poverty trends in china." *Review of Economic Dynamics* **30**, 239–264.

Almås, I. (2012). "International income inequality: Measuring ppp bias by estimating engel curves for food." *American Economic Review* **102**(2), 1093–1117.

Ameriks, J., J. Briggs, A. Caplin, M. D. Shapiro, and C. Tonetti (2020). "Long-term-care utility and late-in-life saving." *Journal of Political Economy* **128**(6), 2375–2451.

Arrow, K. J. (1959). "Rational choice functions and orderings." *Economica* **26**(102), 121–127.

Attanasio, O. and B. Augsburg (2016). "Subjective expectations and income processes in rural india." *Economica* **83**(331), 416–442.

Attanasio, O., B. Augsburg, and R. De Haas (2018). "Microcredit Contracts, Risk Diversification and Loan Take-Up." *Journal of the European Economic Association* **17**(6), 1797–1842.

Attanasio, O., A. Barr, J. C. Cardenas, G. Genicot, and C. Meghir (2012). "Risk pooling, risk preferences, and social networks." *American Economic Journal: Applied Economics* **4**(2), 134–67.

Attanasio, O., R. Bernal, M. Giannola, and M. Nores (2020a). "Child development in the early years." Tech. rep., NBER, WP No 27812.

Attanasio, O., T. Boneva, and C. Rauh (2019a). "Parental beliefs about returns to different types of investments in school children." NBER Working Papers 25513, National Bureau of Economic Research, Inc.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2020b). "Estimating the production function for human capital." *American Economic Review* **110**(1), 48–85.

Attanasio, O., F. Cunha, and P. Jervis (2019b). "Subjective parental beliefs. their measurement and role." Tech. rep., National Bureau of Economic Research.

Attanasio, O. and S. Krutikova (2020). "JEEA-FBBVA LECTURE 2019: Consumption Insurance in Networks with Asymmetric Information: Evidence from Tanzania." *Journal of the European Economic Association* **18**(4), 1589–1618.

Attanasio, O. and V. Lechene (2014). "Efficient responses to targeted cash transfers." *Journal of Political Economy* **122**(1), 178 – 222.

Attanasio, O., C. Meghir, and E. Nix (2020c). "Human capital development and parental investment in india." *The Review of Economic Studies* **87**(6), 2511–2541.

Attanasio, O. and E. Pastorino (2020). "Nonlinear pricing in village economies." *Econometrica* **88**(1), 207–263.

Attanasio, O. P., M. Bornstein, and P. Jervis (2022). "Constructing effective measures of child development: ages 6 to 36 months." Tech. rep., In progress.

Autor, D. H. and D. Dorn (2013). "The growth of low-skill service jobs and the polarization of the us labor market." *American Economic Review* **103**(5), 1553–97.

Banks, J., R. Blundell, and A. Lewbel (1997). "Quadratic engel curves and consumer demand." *Review of Economics and statistics* **79**(4), 527–539.

Ben-Akiva, M. E., D. McFadden, K. Train et al. (2019). *Foundations of stated preference elicitation: Consumer behavior and choice-based conjoint analysis*. Now.

Benhabib, J. and A. Bisin (2011). "Choice and process: Theory ahead of measurement." In "*The Foundations of Positive and Normative Economics: A Hand Book*," (edited by Caplin, A. and A. Schott). Oxford University Press.

Bernheim, B. D., D. Bjorkegren, J. Naecker, and M. Pollmann (2021). "Causal inference from hypothetical evaluations." Tech. rep., NBER, WP No 29616.

Berry, S., J. Levinsohn, and A. Pakes (2004). "Differentiated products demand systems from a combination of micro and macro data: The new car market." *Journal of Political Economy* **112**(1), 68–105.

Bils, M. (2009). "Do Higher Prices for New Goods Reflect Quality Growth or Inflation?*." *The Quarterly Journal of Economics* **124**(2), 637–675.

Bils, M. and P. J. Klenow (2001). "Quantifying quality growth." *American Economic Review* **91**(4), 1006–1030.

Biroli, P., T. Boneva, A. Raja, and C. Rauh (2022). "Parental beliefs about returns to child health investments." *Journal of Econometrics* **231**(1), 33–57. Annals Issue: Subjective Expectations & Probabilities in Economics.

Bisin, A. and T. Verdier (2000). "Beyond the melting pot: cultural transmission, marriage, and the evolution of ethnic and religious traits." *Quarterly Journal of Economics* **115**(3), 955–988.

Black, M. M., K. R. Bromley, V. Cavallera, J. Cuartas, T. Dua, I. Eekhout, G. Fink, M. J. Gladstone, K. L. Hepworth, M. Janus, P. K. Kariger, G. A. Lancaster, D. C. McCoy, G. P. J. McCray, A. Raikes, M. Rubio-Codina, S. van Buuren, M. Waldman, S. P. Walker, and A. M. Weber (2019). "The global scale for early development -gsed." *BMC Public Health* .

Blass, A. A., S. Lach, and C. F. Manski (2010). "Using elicited choice probabilities to estimate random utility models: Preferences for electricity reliability." *International Economic Review* **51**(2), 421–440.

Block, H. D. and J. Marschak (1960). "Random orderings and stochastic theories of response." In "*Contributions to Probability and Statistics*," (edited by Olkin, I., S. Ghurye, W. Hoeffding, W. Madow, and H. Mann). Stanford University Press.

Bloom, N., E. Brynjolfsson, L. Foster, R. Jarmin, M. Patnaik, I. Saporta-Eksten, and J. Van Reenen (2019). "What drives differences in management practices?" *American Economic Review* **109**(5), 1648–83.

Bloom, N. and J. Van Reenen (2007). "Measuring and Explaining Management Practices Across Firms and Countries*." *Quarterly Journal of Economics* **122**(4), 1351–1408.

Bobba, M. and V. Frisancho (2020). "Self-perceptions about academic achievement: Evidence from mexico city." *Journal of Econometrics* .

Boneva, T. and C. Rauh (2018). "Parental beliefs about returns to educational investments—the later the better?" *Journal of the European Economic Association* **16**(6), 1669–1711.

Bornstein, M. H. (1996). "Ideas about parenting in argentina, france, and the united states." *International Journal of Behavioral Development* **19**(2), 347–368.

Bornstein, M. H., W. A. Rothenberg, J. E. Lansford, R. H. Bradley, K. Deater-Deckard, A. Bizzego, and G. Esposito (2021). "Child Development in Low- and Middle-Income Countries." *Pediatrics* **148**(5). E2021053180.

Bourguignon, F., M. Browning, and P. A. Chiappori (2009). "Efficient intra-household allocations and distribution factors: Implications and identification." *The Review of Economic Studies* **76**(2), 503–528.

Browning, M. and P. A. Chiappori (1998). "Efficient intra-household allocations: A general characterization and empirical tests." *Econometrica* **66**(6), 1241–1278.

Browning, M., P.-A. Chiappori, and A. Lewbel (2013). "Estimating consumption economies of scale, adult equivalence scales and household bargaining power." *Review of Economic Studies* **80**, 1267–1303.

Buser, T., M. Niederle, and H. Oosterbeek (2014). "Gender, competitiveness, and career choices." *The quarterly journal of economics* **129**(3), 1409–1447.

Camerer, C. (2011). "The case for mindful economics." In "*The Foundations of Positive and Normative Economics*," (edited by Caplin, A. and A. Schott). Oxford Univ. Press.

Caplin, A. (2021). “Economic data engineering.” Working Paper 29378, National Bureau of Economic Research.

Carson, R. T. (2012). “Contingent valuation: A practical alternative when prices aren’t available.” *Journal of Economic Perspectives* **26**(4), 27–42.

Cavatorta, E. and B. Groom (2020). “Does deterrence change preferences? evidence from a natural experiment.” *European Economic Review* 103456.

Chamberlain, G. and Z. Griliches (1975). “Unobservables with a variance-components structure: Ability, schooling, and the economic success of brothers.” *International Economic Review* 422–449.

Chen, X., L. P. Hansen, and P. G. Hansen (2020). “Robust identification of investor beliefs.” Tech. rep., National Bureau of Economic Research.

Cherchye, L., P.-A. Chiappori, B. De Rock, C. Ringdal, and F. Vermeulen (2021). “Feed the children.” Tech. rep., CEPR Discussion Paper No. DP16482.

Cherchye, L., B. D. Rock, and F. Vermeulen (2011). “The revealed preference approach to collective consumption behaviour: Testing and sharing rule recovery.” *The Review of Economic Studies* **78**(1), 176–198.

Chiappori, P. A. (1988). “Rational household labor supply.” *Econometrica* **56**(1), 63–90.

Christensen, L. R., D. W. Jorgenson, and L. J. Lau (1975). “Transcendental logarithmic utility functions.” *The American Economic Review* **65**(3), 367–383.

Costa, D. (2000). “American living standards, 1888-1994: evidence from consumer expenditures.”

Crawford, I. and J. P. Neary (2021). “New Characteristics and Hedonic Price Index Numbers.” *The Review of Economics and Statistics* 1–49.

Cunha, F., I. Elo, and J. Culhane (2013). “Eliciting maternal beliefs about the technology of skill formation.” *NBER Working Paper* **19144**.

Cunha, F. and J. J. Heckman (2008). “Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation.” *Journal of human resources* **43**(4), 738–782.

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

Curtin, R. (2016). “Katona: a founder of behavioral economics.” In “Handbook of Behavioral Economics,” 30–47. Routledge.

Dardanoni, V., P. Manzini, M. Mariotti, H. Petri, and C. Tyson (2022). "Mixture choice data: revealing preferences and cognition." *Journal of Political Economy* ***forth***.

Deaton, A. and J. Muellbauer (1980). "An almost ideal demand system." *The American economic review* **70**(3), 312–326.

Delavande, A. and B. Zafar (2019). "University choice: The role of expected earnings, nonpecuniary outcomes, and financial constraints." *Journal of Political Economy* **127**(5), 2343–2393.

Deming, D. J. (2017). "The Growing Importance of Social Skills in the Labor Market*." *The Quarterly Journal of Economics* **132**(4), 1593–1640.

Dercon, S. and P. Krishnan (2000). "In sickness and in health: Risk sharing within households in rural ethiopia." *Journal of Political Economy* **108**(4), 688–727.

Dizon-Ross, R. (2019). "Parents' beliefs about their children's academic ability: Implications for educational investments." *American Economic Review* **109**(8), 2728–65.

Dominitz, J. and C. F. Manski (1996). "Eliciting student expectations of the returns to schooling." *The Journal of Human Resources* **31**(1), 1–26.

Dominitz, J. and C. F. Manski (1997). "Using expectations data to study subjective income expectations." *Journal of the American Statistical Association* **92**(439), 855–867.

Doraszelski, U. and J. Jaumandreu (2013). "R&d and productivity: Estimating endogenous productivity." *Review of Economic Studies* **80**(4), 1338–1383.

Doraszelski, U. and J. Jaumandreu (2018). "Measuring the bias of technological change." *Journal of Political Economy* **126**(3), 1027–1084.

Dubois, P., R. Griffith, and M. O'Connell (2020). "How well targeted are soda taxes?" *American Economic Review* **110**(11), 3661–3704.

Dunbar, G. R., A. Lewbel, and K. Pendakur (2013). "Children's resources in collective households." *American Economic Review* **103**(1), 438–71.

Duncan, O. D. (1966). "Path analysis: Sociological examples." *American journal of Sociology* **72**(1), 1–16.

Einav, L., E. S. Leibtag, and A. Nevo (2008). "On the Accuracy of Nielsen Homescan Data." Economic Research Report 56490, US Dept. of Agriculture.

Erdem, T., M. Keane, and T. Öncü (2005). "Learning about computers: An analysis of information search and technology choice." *Quantitative Market and Econonomics* **3**(3), 207–247.

Ertaç, S., A. Hortaçsu, and J. W. Roberts (2011). "Entry into auctions: An experimental analysis." *International Journal of Industrial Organization* **29**(2), 168–178.

Field, E., R. Pande, N. Rigol, S. Schaner, and C. Troyer Moore (2021). "On her own account: How strengthening women's financial control impacts labor supply and gender norms." *American Economic Review* **111**(7), 2342–75.

Gandhi, A., S. Navarro, and D. A. Rivers (2020). "On the identification of gross output production functions." *Journal of Political Economy* **128**(8), 000–000.

Gilbert, M., C. Clark, J. Stone, F. Perroux, D. Lieu, Evelpides, F. Divisia, Tinbergen, Kuznets, Smithies et al. (1949). "The measurement of national wealth: discussion." *Econometrica: Journal of the Econometric Society* 255–272.

Giustinelli, P., C. F. Manski, and F. Molinari (2019). "Precise or imprecise probabilities? evidence from survey response on late-onset dementia." Tech. rep., NBER.

Goldberger, A. S. (1971). "Econometrics and psychometrics: A survey of communalities." *Psychometrika* **36**(2), 83–107.

Goldberger, A. S. (1972). "Structural equation methods in the social sciences." *Econometrica* **40**(6), 979–1001.

Griffith, R. and M. O'Connell (2009). "The use of scanner data for research into nutrition*." *Fiscal Studies* **30**(3-4), 339–365.

Griliches, Z. (1974). "Errors in variables and other unobservables." *Econometrica:* **42**(6), 971–998.

GSED (2021). "The global scale for early development (gsed)." Tech. rep., Early Childhood Matters.

Guiso, L., P. Sapienza, and L. Zingales (2004). "The role of social capital in financial development." *American Economic Review* **94**(3), 526–556.

Guiso, L., P. Sapienza, and L. Zingales (2006). "Does culture affect economic outcomes?" *Journal of Economic Perspectives* **20**(2), 23–48.

Gul, F. and W. Pesendorfer (2011). "The case for mindless economics." In "*The Foundations of Positive and Normative Economics*," (edited by Caplin, A. and A. Schott). Oxford Univ. Press.

Haavelmo, T. (1958). "The role of the econometrician in the advancement of economic theory." *Econometrica* **26**(3), 351–357.

Hamilton, B. W. (2001). "Using engel's law to estimate cpi bias." *American Economic Review* **91**(3), 619–630.

Harris, K. M. and M. P. Keane (1998). "A model of health plan choice:: Inferring preferences and perceptions from a combination of revealed preference and attitudinal data." *Journal of Econometrics* **89**(1-2), 131–157.

Hausman, D. M. (1994). *The philosophy of economics: An anthology*. Cambridge University Press.

Hausman, J. (2012). "Contingent valuation: from dubious to hopeless." *Journal of Economic Perspectives* **26**(4), 43–56.

Heckman, J., R. Pinto, and P. Savelyev (2013). "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes." *American Economic Review* **103**(6), 2052–86.

Heckman, J. J., B. Liu, M. Lu, and J. Zhou (2020). "Treatment effects and the measurement of skills in a prototypical home visiting program." Tech. rep., NBER.

Henrich, J., S. J. Heine, and A. Norenzayan (2010). "The weirdest people in the world?" *Behavioral and Brain Sciences* **33**(2-3), 61–135.

Hu, Y. and S. M. Schennach (2008). "Instrumental variable treatment of nonclassical measurement error models." *Econometrica* **76**(1), 195–216.

Jappelli, T. and L. Pistaferri (2000). "Using subjective income expectations to test for excess sensitivity of consumption to predicted income growth." *European Economic Review* **44**(2), 337–358.

Jayachandran, S., M. Biradavolou, and J. Cooper (2021). "Using machine learning and qualitative interviews to design a five-question survey module for women's agency." Tech. rep., Northwestern University Working Paper.

Jöreskog, K. G. and A. Goldberger (1975). "Estimation of a model with multiple indicators and multiple causes of a single latent variable." *Journal of the American Statistical Association* 631–639.

Jullien, B. (2000). "Participation constraints in adverse selection models." *Journal of Economic Theory* **93**(1), 1–47.

Juster, F., H. Cao, M. Perry, and M. Couper (2006). "The effect of unfolding brackets on the quality of wealth data in hrs." *SSRN Electronic Journal* .

Juster, F. T. (1964). *Anticipations and Purchases: An Analysis of Consumer Behavior*. Princeton University Press.

Juster, F. T. (1966). "Consumer buying intentions and purchase probability: An experiment in survey design." *J. of the American Statistical Association* **61**(315), 658–696.

Juster, F. T. and R. P. Shay (1964). *Consumer Sensitivity to Finance Rates: An Empirical and Analytical Investigation*. NBER.

Kaiser, C. and A. J. Oswald (2022). "The scientific value of numerical measures of human feelings." *PNAS* **119**(42), e2210412119.

Katona, G. (1959). "On the predictive value of consumer intentions and attitudes: A comment." *The Review of Economics and Statistics* 317–317.

Katona, G. (1974). "Understanding consumer attitudes." *Surveys of Consumers* **1976**, 203–219.

Katz, L. F. and K. M. Murphy (1992). "Changes in Relative Wages, 1963–1987: Supply and Demand Factors*." *The Quarterly Journal of Economics* **107**(1), 35–78.

Kaufmann, K. and L. Pistaferri (2009). "Disentangling insurance and information in intertemporal consumption choices." *American Economic Review* **99**(2), 387–92.

Kesternich, I., F. Heiss, D. McFadden, and J. Winter (2013). "Suit the action to the word, the word to the action: Hypothetical choices and real decisions in medicare part d." *Journal of Health Economics* **32**(6), 1313–1324.

Keynes, J. M. (1936). "The general theory of interest, employment and money."

Kuziemko, I., M. I. Norton, E. Saez, and S. Stantcheva (2015). "How elastic are preferences for redistribution?" *American Economic Review* **105**(4), 1478–1508.

Kuznets, S. (1941). "Statistics and economic history." *The Journal of Economic History* **1**(1), 26–41.

Kuznets, S. et al. (1937). "National income and capital formation, 1919-1935." *NBER Books* .

La Ferrara, E. (2019). "Presidential Address: Aspirations, Social Norms, and Development." *Journal of the European Economic Association* **17**(6), 1687–1722.

Lancaster, T. and A. Chesher (1983). "An econometric analysis of reservation wages." *Econometrica* **51**(6), 1661–76.

Lechene, V., K. Pendakur, and A. Wolf (2022). "Ols estimation of the intra-household distribution of expenditure." *Journal of Political Economy* **130**, *forthcoming*.

Lee, J., R. Wedow, A. Okbay, E. Kong, O. Maghzian, M. Zacher, T. Nguyen-Viet, P. Bowers, J. Sidorenko, R. Linnér, M. Fontana, T. Kundu, C. Lee, H. Li, R. Li, R. Royer, P. Timshel, R. Walters, E. Willoughby, and D. Cesarini (2018). "Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals." *Nature Genetics* **50**.

Levinsohn, J. and A. Petrin (2003). "Estimating production functions using inputs to control for unobservables." *The review of economic studies* **70**(2), 317–341.

List, J. A. and C. A. Gallet (2001). "What experimental protocol influence disparities between actual and hypothetical stated values?" *Environmental and Resource Economics* **20**(3), 241–254.

List, J. A., J. Pernaudet, and D. Suskind (2021). "It all starts with beliefs: Addressing the roots of educational inequities by shifting parental beliefs." WP 29394, NBER.

Louviere, J., D. Hensher, and J. Swait (2000). *Stated Choice Models: Analysis and Application*. Cambridge University Press.

Luce, R. and P. Suppes (1965). "Preference, utility, and subjective utility." *Handbook of Mathematical Psychology, III, New York: Wiley* 249–409.

Luce, R. D. (1956). "Semiorders and a theory of utility discrimination." *Econometrica, Journal of the Econometric Society* 178–191.

Luce, R. D. (1959). *Choice Behavior. A Theoretical Analysis*. New York: Wiley.

Luce, R. D. and J. W. Tukey (1964). "Simultaneous conjoint measurement: A new type of fundamental measurement." *Journal of mathematical psychology* **1**(1), 1–27.

Manski, C. F. (1990). "The use of intentions data to predict behavior: A best-case analysis." *Journal of the American Statistical Association* **85**(412), 934–940.

Manski, C. F. (2004). "Measuring expectations." *Econometrica* **72**(5), 1329–1376.

Maskin, E. and J. Riley (1984). "Monopoly with incomplete information." *RAND Journal of Economics* **15**(2), 171–196.

McCoy, D. C., J. Seiden, M. Waldman, and G. Fink (2021). "Measuring early childhood development." *Annals of the New York Academy of Sciences* **1492**(1), 3–10.

Miller, G., Á. de Paula, and C. Valente (2020). "Subjective expectations and demand for contraception." Working Paper 27271, National Bureau of Economic Research.

Miller, S. A. (1988). "Parents' beliefs about children's cognitive development." *Child Development* **59**(2), 259–285.

Mueller, A. and J. Spinnewijn (2021). "Expectations data, labor market and job search." *Handbook Chapter (Draft)* .

Murphy, P. G., James J.and Allen, T. H. Stevens, and D. Weatherhead (2005). "A meta-analysis of hypothetical bias in stated preference valuation." *Environmental and Resource Economics* **30**(3), 313–325.

Neary, J. P. (2004). "Rationalizing the penn world table: True multilateral indices for international comparisons of real income." *American Economic Review* **94**(5), 1411–1428.

Nordhaus, W. D. (1998). "Quality change in price indexes." *Journal of Economic Perspectives* **12**(1), 59–68.

Olley, G. S. and A. Pakes (1992). "The dynamics of productivity in the telecommunications equipment industry." Tech. rep., National Bureau of Economic Research.

Paiella, M. and L. Pistaferri (2017). "Decomposing the Wealth Effect on Consumption." *The Review of Economics and Statistics* **99**(4), 710–721.

Parnes, H. S. (1975). "The national longitudinal surveys: New vistas for labor market research." *The American Economic Review* **65**(2), 244–249.

Pistaferri, L. (2001). "Superior Information, Income Shocks, and the Permanent Income Hypothesis." *The Review of Economics and Statistics* **83**(3), 465–476.

Pistaferri, L. (2003). "Anticipated and unanticipated wage changes, wage risk, and intertemporal labor supply." *Journal of Labor Economics* **21**(3), 729–754.

Plott, C. R. and V. L. Smith (2008). *Handbook of experimental economics results*, vol. 1. Elsevier.

Potter, S., M. Del Negro, G. Topa, and W. Van der Klaauw (2017). "The advantages of probabilistic survey questions." *Review of Economic Analysis* **9**(1), 1–32.

Salz, T. and E. Vespa (2020). "Estimating dynamic games of oligopolistic competition." *The RAND Journal of Economics* **51**(2), 447–469.

Samuelson, P. A. (1938). "A note on the pure theory of consumer's behaviour." *Economica* **5**(17), 61–71.

Samuelson, P. A. (1948). "Consumption theory in terms of revealed preference." *Economica* **15**(60), 243–253.

Schennach, S. M. (2004). "Estimation of nonlinear models with measurement error." *Econometrica* **72**(1), 33–75.

Scur, D., R. Sadun, J. Van Reenen, R. Lemos, and N. Bloom (2021). "The world management survey at 18." *Oxford Review of Economic Policy* **37**(2), 231–258.

Stantcheva, S. (2022). "How to run surveys: A guide to creating your own identifying variation and revealing the invisible." Working Paper 30527, NBER.

Stigler, G. J. and G. S. Becker (1977). "De gustibus non est disputandum." *The American Economic Review* **67**(2), 76–90.

Stone, R. (1954). "Linear expenditure systems and demand analysis: an application to the pattern of british demand." *The Economic Journal* **64**(255), 511–527.

Stone, R. (1984). "Richard stone-prize lecture: The accounts of society." *Nobelprize. org. Nobel Media AB* .

Tobin, J. (1959). "On the predictive value of consumer intentions and attitudes." *The review of economics and statistics* 1–11.

Todd, P. E. and K. I. Wolpin (2003). "On the specification and estimation of the production function for cognitive achievement." *The Economic Journal* **113**(485), F3–F33.

Van der Klaauw, W. and K. I. Wolpin (2008). "Social security and the retirement and savings behavior of low-income households." *J. of Econometrics* **145**(1-2), 21–42.

Visco, I. (1984). *Price Expectations in Rising Inflation*. Contributions to economic analysis. North-Holland.

Wiswall, M. and B. Zafar (2015). "Determinants of college major choice: Identification using an information experiment." *The Review of Economic Studies* **82**(2), 791–824.

Wolpin, K. I. and F. Gonul (1985). "On the use of expectations data in micro surveys: The case of retirement." Tech. rep., Ohio State University.

Wright, S. (1934). "The method of path coefficients." *The annals of mathematical statistics* **5**(3), 161–215.

Appendices

A1 Specification of Measurement Systems

This Appendix provides more details on the general issues in the specification and estimation of measurement systems that were discussed in Section 5.1 of the main text.

One assumption that can be relaxed in the context of the measurement system, and that we mention in the main text, is that of linearity, especially when the available measures are binary or discrete. Here, we explain the logic. In such a situation, a common approach model is the Item Response Theory (IRT). In the case of binary measures, an IRT model relates a factor to a latent measure, which in turn determines the observed outcome. Omitting the superscript for the latent factor θ, we have:

$$m_{it}^{k} = \begin{cases} 1 & if \;\; \alpha_t^k + \beta_t^k \theta_{it} + \varepsilon_{it}^k > 0 \\ 0 & otherwise \end{cases} \tag{A1}$$

The specific IRT model is then determined by assumptions about the distribution of the measurement error term ε_{it}^k; assuming normality, one gets a *Probit* type of model, while assuming a logistic distribution one obtains a logistic relation, or what is usually referred to as a Rasch model. What we are considering is often defined as a 2-parameter Rasch model. Restricting β_t^k to 1, one obtains a 1-parameter Rasch model. In the literature, a 3-parameter Rasch model considers the possibility of random 'correct' answers through an additional parameter.

Another more recent example of IRT use, whose use has become more common in economics, is that of polygenic scores, which aggregate data from many different sites of the human genome and are based on correlations from a wide population data set. Lee et al. (2018), for instance, present estimates of a polygenic score which is associated with individual educational attainment in a specific population. Interestingly, several authors recently noticed that several estimates of the same polygenic score might be available, where the weights to aggregate information from individual loci are based on different samples and/or slightly different methodologies. These alternative estimates can then potentially be used to deal with measurement error problems, as in the model discussed above.

A2 The Tanzania data, new measures and results

The Bukoba region is located in the northwest corner of Tanzania on the western shore of Lake Victoria. The population is around 3 million people which is about 5 percent of the national population. It has an intercensal growth rate 2002/2012 of 3.2 (2.7 average for Tanzania) and the double percentage of children age 6-59 months classified as having malaria, according to the Rapid Diagnostic Test in 2017 in terms of country percentage (15.4% vs 7.3%). The Kagera region is predominantly rural, and agricultural-dependent, producing mainly bananas and coffee in the North and rain-fed annual crops in the South.

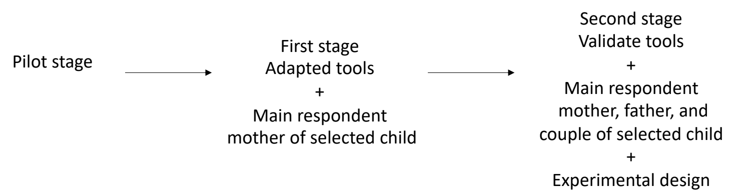
The study was conducted in several stages (see Figure A1). The study sample was randomly chosen to be representative of the total population. The villages were selected from a random sample in Kagera region. In the selected villages, all households were listed, and from those satisfying the selection criteria, a random sample was drawn.

During the pilot period a number of well-known and tested measures, were adapted to the local context using multiple techniques (focus group discussions, ethnography study, and pre-pilot in small samples). We used a multi-method approach for the early childhood development (ECD) instruments, combining three main paths: parent/observer report, direct assessment, and observation (through filming and coding).

After the pilot period, the first stage was implemented in June 2018. Approximately 450 households were recruited across 10 villages. The sample households were randomly selected among those with a child aged between 6 and 36 months and a mother aged between 15 and 25. In this sample, child development was measured using several tools. In addition to standard demographic information about the households involved, the survey also included measures of parental investment.

Information collected during a partial census was used to randomly sample pairs of primary caregivers and children of eligible households from each of the following groups: 150 primary caregivers aged 15-25 years with a child aged 6-11 months, 150 primary caregivers aged 15-25 years with a child aged 12-23 months, and 150 primary caregivers aged 15-25 years with a child aged 24-36 months.

Figure A1: Study Design



In a second stage, which was implemented in August 2018, an additional 450 households were recruited according from the same villages the following procedure. First a random sample of 150 households was selected from a list of 5200 households with children aged 6 to 36 months and mothers aged 15 to 25 years in 8 villages. This is what we will refer to as the mothers sample. In addition, a second list of around 2000 households with children aged 6 to 36 months and the fathers present was identified in 5 different villages.[30] From this list, which also included mothers older than 25, an additional sample of households with fathers present and mothers aged between 15 and 25 years was selected and formed what we label as the couples sample. From the remaining households in the second list, a third sample of 150 households was selected that we label as the fathers sample. Because of the way the second list was formed, the fathers sample comprises couples who are considerably older. Indeed, only 6.6% of the mothers in the sample were younger than 25.[31]

The data collected in the second stage is what we mostly use in this appendix and in Sections 5 and 6. The main goal of the second stage data collection was to validate some of the child development measures constructed with the factor analysis of first stage data. However, we also collected a number of new variables aimed at measuring individual tastes, beliefs, and bargaining power within the couple. Each sample's label corresponds to the respondents in this new set of questions. In the mothers sample, for instance, the questions about tastes, subjective beliefs about parental investment returns, and bargaining power were answered by the mother in private; in the couples

[30]The two sets of villages belonged to different wards, the Kashai and the Bakoba. Both wards belonged to the Bukoba municipality.

[31]In the mothers and couples samples, there are a few mothers older than 25.

sample, the taste questions were answered jointly by the couple while the others by the mother in private; in the fathers sample, fathers answered the bargaining power and taste questions in private. All other (standard) modules were answered by the mother.

All data collection followed a rigorous process of tools, instruments, and survey development.[32] Of the 450 pairs recruited for the last stage of the survey, we obtained usable information from 423 households, comprising 145 in the mothers sample, 136 in fathers sample, and 142 in the couples sample.

The fact that the questions that elicit information on tastes, subjective beliefs about parental investment returns, and bargaining power within the couple are answered by different respondents in the three subsamples makes the survey particularly interesting, as it allows to test the hypothesis that individuals within the family are characterized by differences in these variables. Unfortunately, the recruitment process followed in the field (described above) makes comparisons across different sub-samples difficult to interpret. This is particularly true for the fathers sample, which includes considerably older individuals, as we show below. While the fact that the mothers and couples samples were drawn from different villages is not particularly worrying, given the similarities and the proximity of the villages, the systematic difference in the age structure of the study sample makes comparisons across samples problematic.

In Table A1, we report descriptive statistics on the main features of the three samples discussed above. The table also contains $p-values$ for tests (adjusted for multiple hypothesis testing) of the difference between the mothers and the fathers samples (in column 5) and the mothers and the couples samples (in column 8). In addition to the new measures (discussed below), information on a standard set of variables, including demographics, education, and wealth markers, was also collected in the three samples.

Consistent with the sampling scheme, the mothers and couples samples have relatively younger mothers, with an average age of 22, while in the fathers sample, the average mother's age is about 7 years higher. This difference is also reflected in the average father's age in the fathers sample, (between 5 and 7 years higher than in the mothers and couples samples). About 20% of the mothers sample is made of single

[32] All instruments were translated into the local language, Swahili. Translations were carried out by the field staff following a rigorous back-translation procedure. The household surveys were administered in the child's home by enumerators. The adapted version of the assessments of child development were conducted by trained staff in the presence of the primary caregiver.

Table A1: Household characteristics

	Mothers sample		Couples sample			Fathers sample		
	Mean	Observations	Mean	Observations	Adj. p-value	Mean	Observations	Adj. p-value
Demographics								
Age, M	23.12	145	22.82	142	*0.941*	30.42	136	*0.010*
Age, F	28.98	105	28.83	139	*1.000*	37.05	134	*0.010*
Secondary school or more, M	0.35	144	0.37	142	*1.000*	0.23	136	*0.079*
Secondary school or more, F	0.50	105	0.46	139	*0.990*	0.34	134	*0.059*
Food Share	0.69	145	0.61	142	*0.030*	0.63	136	*0.079*
Wealth index	-0.17	144	-0.04	142	*0.931*	0.22	136	*0.010*
Child characteristics								
Child is male	0.57	145	0.54	142	*0.990*	0.45	136	*0.079*
Age in months	18.44	145	18.25	142	*1.000*	18.36	136	*0.901*
# siblings	0.57	145	0.62	142	*0.990*	2.16	136	*0.010*
Birthrank	2.00	145	2.04	142	*1.000*	3.53	136	*0.010*

Note: The table displays the descriptive statistics for the households."M" indicates values for mothers and "F" indicates values for fathers. Column (5) compares the mothers and couples samples while Column (8) compares the mothers and fathers samples. Columns (5) and (8) reports p-values adjusted for multiple hypothesis testing using the Romano-Wolf approach. Source: Tz Pilot.

mothers (i.e., the child's father is not present). The average child's age is uniform across the three samples and the share of male children is lowest in the fathers sample.

The only significant and somewhat surprising difference between the mothers and couples samples is the share of total expenditure spent on food, which could be considered a useful indicator of economic well-being[33] This variable is higher in the mothers sample, at 0.69,indicating a poorer sample, and lowest in the couples sample, at 0.61.

We also observe a number of wealth indicators, which we use to estimate a wealth index, normalized to have zero mean in the whole sample. Consistent with the evidence on the food share, the wealth index is lowest in the mothers sample. These differences in permanent income and wealth, however, are not consistent with the information on education: both mothers and fathers in the fathers sample are the least educated, while the mothers sample is the most educated. These differences might reflect cohort effects.

A3 Additional Tables

In this section, we present additional tables that are referred to in the main text of the paper.

[33]The intuition for this goes back to the established economic regularity that the food share falls with income Food shares have been used to identify differences in real income and well-being (Hamilton, 2001; Costa, 2000; Almås, 2012; Almås and Johnsen, 2018).

A3.1 Parental beliefs

In this section we present the estimates of the measurement system for parental beliefs. In estimating the perceived effectiveness of parental investment we consider both cognitive, language and socioemotional outcomes. A possible alternative which we have explored is to separate the perceived productivity of parental investment on socio-emotional development from that on cognition and language.

Table A2: Measurement system for Beliefs

	A Measurement System for Beliefs					
	Couples sample		Mothers sample		Fathers sample	
	β	α	β	α	β	α
Language hard	1.000	0.000	1.000	0.000	1.000	0.000
High Dev.	-	-	-	-	-	-
Language hard	2.098	-0.023	1.220	0.024	1.314	-0.042
Low Dev.	*(0.514)*	*(0.201)*	*(0.257)*	*(0.124)*	*(0.243)*	*(0.114)*
Language medium	0.868	0.040	0.573	0.095	1.015	-0.090
High Dev.	*(0.175)*	*(0.074)*	*(0.136)*	*(0.066)*	*(0.138)*	*(0.065)*
Language medium	1.628	0.009	0.556	0.255	1.039	-0.066
Low Dev.	*(0.430)*	*(0.166)*	*(0.172)*	*(0.082)*	*(0.173)*	*(0.081)*
Language easy	0.746	0.065	0.057	0.296	0.877	-0.008
High Dev.	*(0.180)*	*(0.075)*	*(0.091)*	*(0.043)*	*(0.125)*	*(0.058)*
Language easy	1.082	0.167	-0.181	0.556	1.062	-0.075
Low Dev.	*(0.307)*	*(0.122)*	*(0.137)*	*(0.065)*	*(0.179)*	*(0.083)*
Socio-emotional Nine	0.332	0.156	0.566	0.216	0.965	-0.081
High Dev.	*(0.161)*	*(0.067)*	*(0.128)*	*(0.060)*	*(0.169)*	*(0.076)*
Socio-emotional Nine	0.397	0.259	1.420	0.072	1.168	-0.206
Low Dev.	*(0.317)*	*(0.133)*	*(0.238)*	*(0.106)*	*(0.199)*	*(0.090)*
Socio-emotional Five	0.362	0.128	0.379	0.190	0.819	-0.068
High Dev.	*(0.131)*	*(0.055)*	*(0.088)*	*(0.041)*	*(0.135)*	*(0.060)*
Socio-emotional Five	0.024	0.352	0.964	0.118	0.928	-0.182
Low Dev.	*(0.240)*	*(0.100)*	*(0.172)*	*(0.078)*	*(0.170)*	*(0.077)*
Socio-emotional Three	0.425	0.075	0.118	0.191	0.589	0.003
High Dev.	*(0.145)*	*(0.061)*	*(0.065)*	*(0.031)*	*(0.118)*	*(0.053)*
Socio-emotional Three	-0.180	0.374	0.368	0.168	0.702	-0.098
Low Dev.	*(0.202)*	*(0.084)*	*(0.117)*	*(0.055)*	*(0.170)*	*(0.078)*
Factor mean	0.346		0.303		0.348	
Factor variance	0.069		0.151		0.109	

Note: This table shows the loading factors (β) and the intercept (α) for each returns to parental investment on Language and Socio-emotional skills elicited in our survey which are estimated through a measurement system model. Columns (1) and (2) are for the couples sample, Columns (3) and (4) for the mothers sample, and Columns (5) and (6) for the fathers sample. Standard errors in parentheses. Source: Tz Pilot.

A3.2 Relative preferences for children human capital

As explained in the main text, we compute the ratio of (stated) allocation of 4 different commodities allocated to the mother and father to the total expenditure on allocated to children and use these 8 variables as markers of the 'taste for children human capital'. In the Table we report estimates of the measurement system so obtained.

Table A3: Measurement system for relative taste for child human capital

	Relative taste for child human capital					
	Couples sample		Mothers sample		Fathers sample	
	β	α	β	α	β	α
Mother's clothing ratio	1.000	0.000	1.000	0.000	1.000	0.000
	-	-	-	-	-	-
Father's clothing ratio	1.009	-0.021	0.936	-0.039	0.729	0.007
	(0.058)	*(0.013)*	*(0.190)*	*(0.033)*	*(0.165)*	*(0.023)*
Mother's food ratio	0.850	0.043	4.294	-0.527	1.472	-0.031
	(0.172)	*(0.038)*	*(0.649)*	*(0.111)*	*(0.295)*	*(0.041)*
Father's food ratio	0.850	0.001	-0.011	0.111	1.229	-0.023
	(0.150)	*(0.032)*	*(0.179)*	*(0.031)*	*(0.252)*	*(0.036)*
Mother's health ratio	0.540	0.043	-0.028	0.115	1.410	-0.057
	(0.065)	*(0.014)*	*(0.106)*	*(0.018)*	*(0.294)*	*(0.041)*
Father's health ratio	0.530	0.030	-0.005	0.078	0.795	-0.006
	(0.060)	*(0.012)*	*(0.116)*	*(0.020)*	*(0.189)*	*(0.027)*
Mother's transportation ratio	0.540	0.043	-0.028	0.115	1.410	-0.057
	(0.065)	*(0.014)*	*(0.106)*	*(0.018)*	*(0.294)*	*(0.041)*
Father's transportation ratio	0.530	0.030	-0.005	0.078	0.795	-0.006
	(0.060)	*(0.012)*	*(0.116)*	*(0.020)*	*(0.189)*	*(0.027)*
Factor mean (variance)	0.175	(0.020)	0.162	(0.004)	0.135	(0.002)

Note: This table shows the loading factors (β) and the intercept (α) for each ratio of the resources allocated to each spouse for four adult commodities (food, clothing, health, and transportation) from the allocation questions which are estimated through a measurement system model. Standard errors in parentheses. Columns (1) and (2) are for the couples sample, Columns (3) and (4) for the mothers sample, and Columns (5) and (6) for the fathers sample. Source: Tz Pilot.

A3.3 Parental investment

In terms of the measure of parental investment, we use multiple sources of information:

Expenditures for children: Reported spending on children's health, clothing, shoes, toys, education, books and bedding.

Material investments: Reported frequency of acquiring the following products for their child: food, clothing, footwear, confectioneries medicine and oral re-hydration salts, mosquito net, books, toys, and learning materials (notebooks, pens, and pencils).

Play material: The number of toys the child has, made at home or bought, music instruments, books, and drawing equipment.

Adult activities with children: Reading books, singing, playing, and cooking with the child.

Didactic scale: From the Parental Style Questionnaire (PSQ): whether the primary caregiver (i) spends time playing with the child, (ii) provides the child with independent time to explore and learn on his/her own, (iii) provides the child with diverse social and interactive experience with same-age peers through play groups or informal get-together, (iv) provides the child with a structured organized, and predictable environment, (v) provides language learning opportunities for the child by labeling and describing qualities of objects, events or activities, reading books etc., (vi) provides the child with a variety of toys and objects for play and exploration, (vii) is patient with the child's misbehavior, and (viii) is flexible about behaviors expected from the child.

Social scale: The social scale from the PSQ accounts for whether the primary caregiver (i) promptly and appropriately respond to the child's expressed distress or discomfort, (ii) spend time talking to or conversing with child, (iii) provide child with quick and positive feedback to his/her bid for attention, (iv) provide child with affectionate displays of warmth and attention, and (iv) is aware of what child wants or feels.

Table A4 reports descriptive statistics on the components of parental investment.

Table A4: Parental Investment Components

	Couples sample	
	Mean	Observations
Parental Investment		
Raw Activity Score (/26)	14.89	142
Raw Material Investment Score (/12)	5.23	142
Raw Play Material Score (/8)	1.51	142
Social scale (/5)	4.28	142
Didactic scale (/8)	4.94	142
Expenditure on children share	0.32	142
Total Parental Investment (factor)	-0.48	142

Note: Mean and number of observations for markers of parental investment for the couples sample. In parenthesis the maximum value of each marker of parental investment. Source: Tz Pilot.