



**AgEcon** SEARCH  
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

*The World's Largest Open Access Agricultural & Applied Economics Digital Library*

**This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.**

**Help ensure our sustainability.**

Give to AgEcon Search

AgEcon Search  
<http://ageconsearch.umn.edu>  
[aesearch@umn.edu](mailto:aesearch@umn.edu)

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

# The Econometrics of Hope

Nargiza Chorlieva, University of Alberta, [chorieva@ualberta.ca](mailto:chorieva@ualberta.ca)

Brent Swallow, University of Alberta, [bswallow@ualberta.ca](mailto:bswallow@ualberta.ca)

Sandeep Mohapatra, University of Alberta, [smohapat@ualberta.ca](mailto:smohapat@ualberta.ca)

***Selected Paper prepared for presentation at the 2022 Agricultural & Applied Economics  
Association Annual Meeting, Anaheim, CA; July 31-August***

*Copyright 2022 by [Nargiza Chorlieva, Brent Swallow, Sandeep Mohapatra]. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.*

# Abstract

We rarely think of hope as a determinant of economic outcomes. Yet it has been shown to be a great motivator of human behaviour in the fields of psychology and anthropology. Even though economists started to pay considerable attention to hope as a potential solution to poverty and economic development, economics literature still does not pay immense attention to measurement of hope, while psychology does. The objective of this study is to apply the advanced measurement techniques to the discussion of hope in the context of economics. Our first objective is to ascertain the efficacy of standard survey measures of hope used commonly in the literature on the economics of hope. Our second objective is to determine if the hope survey questions consider heterogeneity between individuals belonging to different demographic characteristics including gender, and access to public services. To meet our objectives we rely on a set of primary data collected in two rural regions of Tanzania with a wealth of mindset questions (hope, aspirations, self-efficacy, social cohesion etc). With regards to our first objective, which we address using Item Response Theory (IRT) methods, we find clear evidence that respondents perceive positive and negative questions differently. The former perform well at discriminating people with low levels of hope, while the latter are useful for identifying individuals with high levels of hope. With regards, to our second objective, which we address using Differential Item Functioning (DIF) methods, we find significant evidence of heterogeneity. These findings lead to a greater understanding of the psychological differences that underpin subgroup responses and allow us to consider them while making further causal analyses.

# 1.Introduction

In recent years, the psychological concept of hope has received increasing attention from economists, leading to the emergence of a new field that blends psychology, economics, and anthropology. Empirical evidence shows that hope and aspirations affect economic behavior such as individuals' motivations for savings and investments in their futures through entrepreneurship, education, health, or working with others to solve social issues (Banerjee et al. 2015; Duflo 2012; Lybbert and Wydick 2018; Lybbert et al. 2016).

An economic approach to the theory of hope considers how our attitudes toward the future affect our current behavior. According to an emerging literature on the economics of hope (Lybbert and Wydick 2018), an individual's psychological attributes, including hope and aspiration levels create "internal constraints" that play a significant role in welfare and poverty outcomes. This constraints with standard approaches in economics which typically more focused on how to relieve individuals' "external constraints" such as a lack of access to schools, hospitals, roads and markets (Besley 2017; Glewwe, Ross, and Wydick 2014).

A number of studies have demonstrated that additional interventions addressing such internal constraints significantly complement the results obtained from interventions that exclusively target external constraints (Banerjee et al. 2015; Bloem and Wydick 2021). One example is a study conducted in the Philippines where children

who attended a faith-based kindergarten program had a higher level of academic performance compared to children in a government-sponsored kindergarten program (Bloem and Wydick 2021). According to the authors, the difference may be explained by the mediating effects of socio-emotional skills such as self-control, openness and conscientiousness among the faith-based kindergarten group. Thus, in this example an intervention targeting internal constraints, i.e. the faith-based program, significantly complemented the intervention addressing the external constraints, i.e. the standard education program.

At the same time, recent field experiments show that attitudes towards hopelessness and risk aversion pose obstacles to growth and human development in many developing countries. Duflo (2012) argues that providing grounds for hope helps break the “poverty trap” - a set of self-reinforcing mechanisms where a current poverty is a direct cause of future poverty. Duflo (2012) argues that many of the very poor may avoid investing their limited resources in their futures because they believe such efforts are unlikely to yield significant benefits. This belief can make it rational for those living in extreme poverty to constantly prioritize immediate consumption above things like education, which could possibly help them get ahead and out of poverty. A lack of hope for a better life appears to induce even rational agents to relinquish possibilities to improve their living standards, and thus to stay mired in poverty.

Although these findings signal the importance of hope in the economics of poverty, several important caveats exist. First, psychological constructs like hope are not directly observable (i.e they are latent) and do not readily lend themselves to empirical analysis. It is difficult for economists to conduct experiments, create theories, or enhance interventions without the ability to appropriately measure the intended constructs. Banerjee2015 (2015) presents results from six randomized control trials of an integrated approach to improve livelihoods among the very poor. In addition to transferring a productive asset, the participants also receive services such as consumption support, training and coaching. The approach resulted in the increase in

self-employment income, as well as an improvement in psychological well-being of the participants. The authors admit that the improvements of psychological well-being cannot be fully grasped by the existing economic theories and a more detailed psychological measurement is necessary to understand the underlying mechanism. Second, a systematic review on measuring hope shows that most studies were conducted in developed countries (Redlich-Amirav et al. 2018), which may not translate well to the context of developing countries. Therefore, our work can shed more light on the applicability of the hope measurement to a different context. The third caveat is that two-thirds of the studies on the measurement of hope were conducted in educational settings or in health facilities. Since these are very specific samples, the results may not be transferable to the general population (Redlich-Amirav et al. 2018). Our data is not limited to any specific population category but still constrained to rural Tanzania. Our study can provide more insight into how the hope scale performs in the general population living in rural areas of developing countries. Fourth, measuring hope can be challenging because of measurement invariance. Measurement invariance is an underlying assumption in psychometric measurements tools that the items under consideration are understood in the same way by different populations. Understanding individual and group differences in measuring hope levels may be critical given their apparent influence on the poor’s economic behaviour and outcomes. In addition, the interpersonal comparability of a quantitative measure of hope is unexplored and requires investigation (Bloem et al. 2018). The issue of interpersonal comparability arises when survey respondents understand the same concept in vastly different ways. Given these problems in the measurement of hope, psychometric measurement tools, particularly short and concise surveys are critical to adequately assessing, planning interventions, and tracking development activities in the region. Development agencies seem to be recognizing this fact as they are paying increasing attention to hope and a range of other latent variables. Given good psychometrics, the hope survey should provide significant value to the mindset measurement tools.

### 1.0.1 Objectives

Taken together, the overall goal of this work is to contribute to the emerging literature on hope through addressing these caveats via two objectives. To do so we ask the following questions:

1. Are the survey questions measuring hope adequately?
2. Are different subgroups (e.g. men and women) perceiving the questions the same way?

To meet our objectives, our study team carried out a hope survey with more than 5200 individuals as part of a comprehensive household survey collected from 2816 households living in 98 villages in Kigoma and Iringa region, Tanzania. This is a preliminary stage of future field experiments designed to measure the economic effects of interventions aimed at changing people’s intrinsic motivation in the region. The hope survey we use was adapted from the Scioli Hope Scale which defines hope as “a future-directed, four-channel emotion network comprising four constituent channels of mastery, attachment, survival, and spiritual systems” Scioli et al. The short version of survey we use has 12 items, of which 8 are positively worded and 4 are negatively worded. They were carefully translated into Swahili. Respondents are asked to rate the extent of agreement with these items on a 4-point Likert-type scale ranging from strongly disagree, disagree, agree and strongly agree.

We answer the first question in Section 4 where we applied Item Response Theories (IRT). Specifically, we focused on Graded Response Model (GRM) using IRT. Our aim is to see whether questions in our survey measure hope as they are supposed to.

We answer the second question in Section 5 where we used differential item functioning (DIF) analysis, which helps to meaningfully compare different sub-groups in terms of their tendencies to respond in certain ways to the survey items. Our goal in this section is to determine if hope questions consider discrepancies between people of different demographic characteristics and their access to public services.

## 2. Background and Literature Review

The objective of this section is to situate our study in the existing economic literature of hope. To do so, we first review the empirical work that has analyzed hope and aspirations and how it might change behaviour. We then review the hope literature for developing countries, as well as discuss the distribution of hope across different groups of people. In the background section we discuss different definitions of hope, review the literature on how latent traits like hope have been measured, provide a brief overview of Tanzania and the World Vision. We use “latent trait” and “latent construct” interchangeably hereafter.

### 2.1 Literature Review

#### 2.1.1 The Economics of Hope

Anthropologist Appadurai (2004) states that future orientation, an important aspect of hope, is a fundamental and hidden determinant of behavior which has radical implications for poverty alleviation and development. He views hope as a “navigational capacity”, which is like a map that shows how to get to where you want to go in the future (Appadurai 2004). Following Appadurai (2004), Ray (2006) analyzed how poverty and aspiration or hope failures may be mutually reinforcing and built models that have been critical in the study of aspirations. Individuals compare themselves to their peers as well as those within their grasp (Ray 2006). Aspirations are shaped

by the lives and accomplishments of “similar” or “attainable” individuals, or in other words, of a “local society”. A related work by Snyder (2002) maintains that a key part of hope is that someone can see a clear link between what they do now and what they will do in the future. In the absence of this element, even if the person has relatively low aspirations and a lot of agency to accomplish it, he or she may not always behave in a future-oriented manner. Furthermore, true agency and pathways may not always be the same as perceived agency and pathways. (Snyder et al. 1991; Snyder 2002).

Following Appadurai (2004) and Ray (2006), and concepts of hope in psychology, development economists Lybbert et al. (2016) developed an economic model of hope. The framework is based on a reference-dependent utility that incorporates aspirations, agency, and pathways. It demonstrates how hope and ambition can influence internal development limits and potential poverty traps. They examine and differentiate “aspirational hope” from “wishful hope,” arguing that whereas aspirational hope – “hope to...” is associated with growth and productivity, wishful hope – “hope that...” is an optimistic form of hope that results in disappointment and risky behaviour. This conceptual framework is valuable in quantifying hope in poorer nations, since the poor may lack agency and rely on outside assistance. Nevertheless, studies on the role of hope in poverty alleviation are mostly theoretical and lack empirical support (Duflo 2012). Lybbert et al. (2016) also talk on how low self-efficacy and internalised constraints might make people act a certain way. For example, a young girl thinks that jobs as engineers are not accessible to women, so she puts less effort into school. This internalisation of constraints on pathways is different, however, from a scenario where low self-efficacy makes the girl think she can not keep up with the grades she needs to get a degree. This could be a poverty trap since the effort required to determine what might be true limits lies beyond the equilibrium path. Sen (2014), also talks about internal constraints as a possible reason why a person might not be able to do what they want. People think the constraints on their way out of poverty are

more restrictive than they really are.

### **2.1.2 Hope and Behaviour in developing countries**

The role of hope can be useful especially in developing countries because of conditions created by a variety of conflicts, disasters, social prejudices. Previous work shows that hope is a powerful motivator of human behaviour in such contexts (Kleist and Jansen 2016; Luthans and Jensen 2002; Snyder et al. 2000; Webb 2007). It is associated with less stress, higher life satisfaction, well-being, and improved quality of life. It is positively correlated with resilience, self-efficacy, and is negatively correlated with depression and anxiety (Hutz et al. 2014; Rustøen, Cooper, and Miaskowski 2011; Snyder 2002). Becoming a non-smoker, eating more fruits and vegetables and engaging in regular physical activity are examples of health-enhancing actions that have been correlated with hope (Berg et al. 2011; Nothwehr, Clark, and Perkins 2013; Anderson and Feldman 2020). Hope has been linked to mental health issues, such as anxiety, depression, suicide, substance abuse, stress, post-traumatic stress and well-being (Gallagher and Lopez 2009), as well as academic achievement (Snyder 2002), productivity and leadership in the workplace (Adams and Pulvers 2002)

When it comes to empirical research of hope and aspirations in development economics, they were tested primarily in the domain of randomised control trials (RCTs). The RCT is a type of study in which people are randomly assigned to one of two groups: the experimental group, which gets the treatment that is being tested, and the control group, which gets an alternative treatment that is not being tested. The two groups are then followed up to discover if their outcomes differ. The trial's results and analysis are used to assess the intervention's effectiveness, or how much a treatment, procedure, or service benefits patients. RCTs are the most rigorous method for establishing a causal relationship between the intervention and the outcome (Sibbald and Roland 1994). Beaman et al. (2012) investigated whether observing female political leaders raises aspirations of girls and their parents by using a self-reported set

of five questions on educational attainment, future occupation, age of marriage, and leadership potential. By exploiting RCTs in West Bengal they found that the gender gap in parents' and adolescents' aspirations decreases due to a role model effect in villages where women were assigned to leadership positions. They say that aggregating aspirational questions in one index may not be legitimate because, as the questions may belong to different spheres of aspirations, they may confound each other's effects. Thus, after some validation of the survey questions they excluded one question with the lowest correlation from the aspirations index (Beaman et al. 2012). An earlier study conducted by Jensen and Oster (2009) using individual panel data found that exposure to cable television affects the aspirations of women by lowering the tolerance towards domestic violence, higher school enrollment of children, increased women's autonomy and decreased fertility among adult women. This happened not because television caused building a school or reduced school fees, but television exposed rural households to urban lifestyles, values, and behaviors that are very different from their own (Jensen and Oster 2009). Adding to their study that such interventions can be effective in shifting perceptions and behaviors, Tanguy et al. (2014) carried out a field experiment in Ethiopia where poor people listened to stories of how people similar to them escaped poverty. Six months later they found that the treatment group had experienced significant positive impacts on aspirations index, actual savings, investment in education, and time allocation for business. They argue that the intervention is not linked to updating beliefs because of new information, but rather a "vicarious experience" of how someone similar to them could get through poverty due to hard work and persistence and this inspires them to have more hope in their children (Tanguy et al. 2014). These results support the hypothesis set by Appadurai (2004) and Ray (2006) that aspirations, although an individual attribute, respond to collective influence. Despite the huge economic and psychological barriers that the very poor face, even relatively minor changes in their circumstances can considerably raise their chances of escaping poverty. For example, providing a single asset (a cow or

a sewing machine) to very poor people in West Bengal resulted in a sustained increase in consumption that exceeded the asset value, and the beneficiaries increased their overall income in ways not entirely traceable to the initial assets (Duflo 2012). The transfer's effects on economic behaviour and emotional well-being much surpassed what the researchers could have anticipated based on the transfer's economic value alone.

In behavioural economics, a growing body of studies examines the relationship between self-esteem and economic outcomes. For example, Bénabou and Tirole (2003) demonstrate that empowering and encouraging an individual can increase self-esteem, which can result in increased achievement (Bénabou et al. n.d.). Darolia and Wydick (2011) discover that measures such as parental praise aimed at increasing self-esteem result in academic accomplishment above what inherent ability would indicate for university students and the more self-esteem a child has, the more likely he or she is to try harder, believing that effort put into a task will pay off in the long run. students who received gifts of large spending allowances and cars showed lower levels of effort and academic achievement (Darolia and Wydick 2011). Other related work by Duggal et al. (2016) associates hope with a positive mood that has direct impacts on human health (Duggal, Sacks-Zimmerman, and Liberta 2016). Genc and Arslan (2021) link hope to subjective well-being, a concept that is gaining increasing recognition in economic studies. According to the findings of a study conducted on college students in Turkey, young adults who experience high levels of stress because of the coronavirus have less optimism and hope, resulting in lower subjective well-being (Genc and Arslan 2021). Marlon et al. (2019) claim that both hope and doubt are strong determinants of climate change mobilization via political behaviour and support for greenhouse gas mitigation policies. They argue that hope is not always good, and doubt is not always bad; the mix of constructive hope and doubt can be motivating, whereas false hope (e.g., wishful thinking) and fatalistic doubt (e.g., beliefs that there is nothing humans can do) can lead to avoidance, distancing, and inaction (Marlon

et al. 2019).

### **2.1.3 Is hope same for all?**

Hope is not evenly distributed among different groups. It is affected by a person's position in society, exposure to other practices and ways of life, access to and control over resources, and, most crucially, knowledge. For example, some women in Bangladesh aspire to eat two meals a day, to look clean and wear good clothes. They do not use loans available, because they do not have the capacity to start a business or are willing to take risks. But they have ambitious aspirations for their children, they want them to be educated. On the other hand, the wives of pond owners and other market operators have high aspirations not only for their children, but for themselves as well. They want to accumulate and run their businesses on a bigger scale, take ponds on lease, use capital to hire other people, build a cement-brick house, buy a TV, and so on (Nathan, 2005). Aspirations among women and men can also be different due to gender division of responsibilities, their economic roles in the household, and cultural practices. Another example, the poor don't have the opportunities or routes they need to reach their goals, because they have lower chances of seeing and practicing that navigational capacity. Their aspirations are drawn from their own experience and are strongly affected by their social environment (Appadurai, 2004).

Snyder (1994) developed a scale for adults and children's hope in order to determine whether there are differences in hope scores by ethnic group, gender, or age. This aided in the development of interventions tailored to specific groups. For example, despite efforts to equalise samples based on socioeconomic background, the study in hope between public and Catholic schools discovered considerable differences in hope between schools. Separate examinations of the samples revealed apparent racial distinctions (McDermott et al., 2002). De Sá et al., (2019) examined gender and age variations in depression measurement among Brazilian college students using item response theory and differential item functioning tests and discovered that women

and younger students are more likely to suffer from depression than older and males. However, results indicated that statements such as “crying” cause women to overestimate their symptoms and older students to indicate “loss of interest in sex”. As such, it may be critical that measures used to assess hope incorporate gender and age differences.

## **2.2 Background**

### **2.2.1 Roots, definitions and hope in other disciplines**

We see that each scientific field takes a unique approach to the measurement of hope. For example, the most widely researched model of hope in the psychology literature is Snyder’s (2002) Hope Theory. According to Snyder’s definition, someone who has hope has a goal in mind, sees a way to get there, and thinks he/she has the agency (power) to get there (Snyder et al. 2000). However, Snyder’s theory has been criticized for being overly individualistic (Du and King 2013), for focusing too much on personal control and action and too little on trust (Tennen and Tennen 2002), for failing to distinguish enough between hope and optimism (Aspinwall and Leaf 2002), for ignoring the significance of emotion and for departing significantly from how individuals perceive hope in daily life (Tong et al., 2010). As a result of these and other criticisms, a few more theories have been developed that focus on spiritual, social, open-ended hope and multidimensionality of hope (Scioli et al. 2011; Du and King 2013).

In this paper, we use Scioli et al.’s (2011) definition of hope. Scioli et al. (2011) created an alternative model of hope as a future-directed network of emotions using literature from psychology, philosophy, theology, spirituality, and nursing. They explain hope in terms of a “network” of four interacting components collaborate such as mastery, attachment, survival, and spirituality. It is broad and integrative, going beyond goals to include social support, terror management, and life meaning (Scioli

et al. 2011). Given that hope can be understood in these different ways, and analyzed from different viewpoints, we emphasize our analysis of the Scioli et al (2011) conception of hope is one of several possibilities.

### **2.2.2 Brief overview of Tanzania**

Tanzania, East Africa’s second-largest country, has one of the region’s lowest population densities. It is a resource rich country that has a great potential in the agricultural and mining industries. (*World Bank*, 2021). Nonetheless, according to the multidimensional poverty index from 2015/2016 data, which identifies multiple overlapping deprivations suffered by individuals in 3 dimensions: health, education and standard of living, 55.4 % of the population in Tanzania was classified as multidimensionally poor. Tanzanians wanting to overcome poverty face an uphill battle due to a lack of employment opportunities, especially in rural areas. In 2019, 15% of Tanzanian youngsters were neither working nor enrolled in education (*Faith and Development in Focus: Tanzania*, 2018).

Kigoma is a region in northern Tanzania, located on the shores of Lake Tanganyika. It borders Burundi to the north and the Congo to the west. Iringa is bounded to the north by Singida and Dodoma, to the east by Morogoro, to the west by Mbeya, and to the south by Njombe. The Kigoma region has a long history of hosting refugees from Burundi and Congo. The majority of the refugees hosted are Burundians who fled their country in 2015.

In Tanzania, undernutrition remains one of the country’s biggest human development challenges. (IRIS 2017). According to estimates, 450,000 children in Tanzania are acutely malnourished or wasted, with 100,000 suffering from the most severe form of malnourishment. Nationally, chronic malnutrition affects 34.7% of under-fives, while 11.5% suffer from severe stunting. In Kigoma chronic malnutrition exceeds 40%, while in Iringa more than half the children are chronically malnourished (51.3%) (*IRIS*, 2017).

In 2019, Iringa had greater enrolment rates than the national average, while Kigoma had lower enrolment rates. Kigoma also has a gender disparity that favours boys. (*The World Bank Group. Tanzania Gender Assessment* 2022). In Kigoma farming as an employment status accounted for 78% while in Iringa 70.3%. The issue on ownership of certain assets builds poverty monitoring indicators. The percentage of households who own assets is about twice as high in Iringa as it is in Kigoma. (*National Bureau of Statistics* 2014)

Currently, 1.6 million Tanzanians—or 4.6% of the population are living with HIV, 72,000 of whom were newly infected in 2018. Prevalence rates vary widely by region, HIV prevalence rate in the province of Iringa is up to 11.3% among adults aged 15 years and older, while in Kigoma is 3% (*Tanzania HIV Impact Survey 2016-2017: Final Report*, 2018). Women are more likely to get infected than men because they marry earlier, have older partners, and have more trouble negotiating safe sex. In Tanzania, 31% of girls get married before they turn 18. This makes Tanzania the eleventh country in the world with the most child brides. Interesting that Iringa, one of our sites has the lowest percent, about 8% child brides in the country (*Tanzania, Girls Not Brides*, 2021).

## 3. Conceptual Framework

In this section we will review the empirical methods we will use to meet our two objectives. First, we will provide a summary of Item Response Theory (IRT), and its assumptions. Then we will introduce the specification of IRT called Graded Response Model (GRM) which we will use to meet our first objective. Finally, we will introduce a method called Differential Response Functioning (DIF), which will be used to meet our second objective.

### 3.1 Measurement of latent constructs

Hope is complex. Hope, aspirations, self-efficacy and social cohesion are latent constructs that have interpretation issues which can be measure either by observing behaviour or asking subjective questions. They are limited to subjective assessment like depression and anxiety that are internal to a person and interact with her external constraints to produce outcomes. Thus, borrowing from psychology this process has been harnessed into the domain of economics. Rigorous instruments of measuring hope used in the psychology literature may not only help to predict future economic behavior of people but also draw a distinction between wishful and aspirational hope discussed before.

A good measurement instrument of latent constructs includes a list of one or more questions or statements which collectively result in the accurate assessment of the

latent trait. In psychometrics, these statements or questions are referred to as “items”. In psychometric testing, item construction is an essential component (Furr 2014). Items should also be worded in a way that doesn’t lead respondents to answer in a certain way. Items should also refer to a single concept or issue to ensure clarity. (*Core Guide: Measure Development*, 2020).

## 3.2 Item Response Theory

The method we are using to meet our first objective is called Item Response Theory (IRT). IRT is a probabilistic model of how a respondent would respond to any given item(s) depending on their degree of a latent trait. In our case, an individual’s hope level (latent trait denoted by  $\theta$ ) is the probability of endorsing the item. So, the more an individual’s hope, the greater the probability of providing an answer “strongly agree” to positive items and “strongly disagree” to negative items. The probability of endorsing them increases monotonically as the respondent’s hope level increases (Embretson 2000). The two-parameter logistic (2PL) model is a frequently used IRT model for dichotomous items (Birnbaum, A. 1968). This function depends on two parameters - the item’s discrimination ( $a$ ) and difficulty ( $b$ ). Difficulty is the minimum latent trait level required to correctly answer an item. It is also known as “location” for their location on the difficulty range. Discrimination (slope) is an item’s ability to distinguish between individuals with low and high latent trait levels. We may build an item characteristic curve (ICC) based on the item’s difficulty and discrimination to determine a respondent’s chance of endorsing the item in relation to his or her latent trait level.

In IRT, the latent trait and item difficulties are typically rated on a standardized metric across a sample of respondents, with means of 0 and standard deviations of 1. For example, a person with a hope level of 0 has an average level of hope, and a person that has a hope level equal to 1 is 1 standard deviation above the average. An item with a difficulty level equal to 0 has an average difficulty, while an item with a

difficulty level of 1 is a relatively difficult item. If an item has a difficulty of 0, then a person with an average level of hope will have a 50% chance of agreeing with the item. If the discrimination is positive and high, then the item is consistent with the trait we are measuring. Low levels of discrimination indicate that not only hope, in our case, is needed to positively endorse the item, but also something else. If the discrimination is 0, then it is unrelated to the trait level. And if it is negative, then it is inversely related to hope, meaning that high levels of hope make respondents less likely to agree with the items. A perfect item would be highly discriminating and have threshold settings that span a wide range of the latent variable (Embretson 2000).

Ideally, items should provide strong information across the whole spectrum of the latent trait. In the worst case, all our items give little information for any of the hope levels. We also want to avoid a situation when our items' informativeness is concentrated around same levels of hope. When the IRT model provides information functions for each item, it is called an item information function (IIF), which shows the reliability of an item at different points of the trait. Items need to capture a range of hope levels (low, medium, or high). Items may differ in two respects. For example, a very easy math question will provide little information on who has high ability since most will answer it correctly; similarly, a very difficult question will carry little information about which individuals have low ability because it is too difficult for them.

Information functions for all items together is called test information function (TIF). It provides the reliability of all items together. IRT additionally offers the test information function's standard error. The SE of a test information function is the inverse of the test information function; therefore, both are measures of the test's precision at various trait levels. IRT also shows the standard error (SE) of the TIF, which indicates the precision of the test at different trait levels (Embretson 2000).

The proper use of IRT requires a set of assumptions to hold. The first assumption

is monotonicity which means that the probability of a person endorsing an item increases as the person's latent trait level increases. The second IRT assumption is uni-dimensionality, which refers to the notion that the items on a scale all measure the same thing: that they are all impacted by a single latent feature. The third IRT assumption is invariance. If the invariance assumption is satisfied, respondents with similar latent traits will get similar scores on the same or a completely different test designed to measure the same trait. In addition, respondents who have different latent traits will keep the same difference in their scores if they take the same test. The fourth and final IRT assumption is local independence. This hypothesis refers to the responses of subjects with the same ability or trait level on different test items being independent of each other (Reese 1999).

### 3.3 Graded Response Model

An important step in an IRT analysis is choosing an appropriate measurement model. Polytomous IRT models are required to reflect the nonlinear relationship between a respondent's latent trait and the likelihood of responding in each category for multiple-category item-response data. When item responses are coded into more than two ordered categories, Samejima's graded response model (GRM) is appropriate (Samejima 1969). The GRM is specified as follows:

$$P(X_{ri} \geq k | a_i, b_{ik}, \theta_r) = 1 / (1 + \exp(-a_i(\theta_r - b_{ik}))) \quad (3.1)$$

where  $X$  represents respondents  $r$ 's response to each item  $i$ . The GRM models the probability of a respondent selecting a score at or above each item score category. With a four-category Likert scaled item scored from 1 to 4, the probability of scoring at or above 1 is 1, but for score points  $k = 2 \dots 4$ , the probability is modeled as four logistic functions rising from 0 to 1 across the latent score scale,  $\theta_r$ . Here,  $\theta_r$  represents the latent score for each respondent  $r$  (hypothesized to indicate a respondent's

level of hope),  $a_i$  is the information parameter for each item  $i$ , and  $b_{ik}$  represents the location parameter for each item  $i$  and score category  $b_{ik}$ . The information parameter ( $a_i$ ) indicates how well an item can distinguish between respondents with very similar hope levels. The location parameter ( $b_{ik}$ ) indicates whether respondents need a higher or lower level of hope,  $\theta_r$ , to respond at or above that level  $k$ . The  $\theta_r$  parameters can be interpreted on a standard normal scale, where 1 and +1 are one standard deviation below and above the mean, correspondingly. In a strict sense,  $a_i$  is the increase in the log of the odds of scoring at or above item  $i$ 's categories  $k$ , for each 1-standard-deviation-unit increase in  $\theta_r$ . The location parameters,  $b_{ik}$ , indicate where a respondent with  $\theta_r = b_{ik}$  has a 50% chance of scoring at or above category  $k$  of item  $i$ . These parameters are invariant between item and respondent populations. Figure 18 in the appendix illustrates histograms at the item level. It shows whether each response category has enough respondents replies to determine the appropriate location parameters  $b_{ik}$ .

### 3.3.1 Model choice and model fit

When designing our IRT study, we can assume that all of the items are equally effective at discriminating between respondents, or we can assume that each item has a unique discrimination parameter. Thus, we are comparing a “constrained” model to an “unconstrained” model - and, as we will see, the difference between these models in terms of their fit to the data can be explicitly tested. To compare the model fit statistics of constrained and unconstrained models, we use an ANOVA (Analysis of Variance) that evaluates the difference between the two models.

## 3.4 Differential Item Functioning

The second objective of this paper is to find out if the hope questions are understood the same way by individuals of different backgrounds based on gender, age, region, religion, food security, access to water and treatment groups. To meet the second

goal, we will use Differential Item Functioning (DIF) analysis, which will help to meaningfully compare different sub-groups. Following the estimation of IRT item attributes, we can see if the items' properties (e.g., item discrimination and difficulty) are invariant across groups. In item analysis, IRT is a helpful analytical technique for detecting group differential item functioning (DIF) (Thissen 1993). Researchers use DIF analysis to investigate the potential that an item may operate differently for one focal group (e.g., females) than for another reference group (e.g., males) despite adjusting for variations in the assessed construct between the two groups. Typically, the reference group refers to a group presumed to be in an advantageous position, while the focal group is suspected to be at a disadvantage. When an item has a DIF, people with the same characteristic level but belonging to different groups (e.g., males versus females) have a varied chance of responding to the item. DIF items are sometimes referred to as "biased items" since they lead one group to score higher on a measure than the other merely due to their group membership.

DIF can happen for several reasons, including different interpretations, group norms on how to respond (e.g., avoiding extreme categories, such as "strongly agree" or "strongly disagree"), or other things such as how items are administered. As a result, assessing for DIF can be very useful in finding items that show cultural bias, which can then be changed or removed. Finding DIF is generally undesirable since such items might risk the validity of an instrument used to assess the trait levels of individuals from various populations or groups. On the other hand, DIF may lead to a greater understanding of the psychological differences that underpin subgroup responses, as well as future study into mechanisms for such differences. The presence of a DIF implies a conditional difference in item performance across groups, but it does not reveal the source of the DIF. Understanding the cause of DIF is important for two reasons: (a) it helps test developers better understand item and group properties that are responsible for differential test performance, and thus whether the DIF represents a fair or unfair advantage on an item; and (b) it can help test developers

better understand item and group properties that are responsible for differential test performance, and thus whether the DIF signifies a fair or unfair advantage on an item.

For example, a depression index has a question about the frequency of crying (De Sá et al. 2019). However, if women endorse this item more than men, keeping the level of depression constant, there is DIF. Therefore, the depression index will be biased towards diagnosing women more frequently than men if the item containing a DIF is not addressed. We will conduct a DIF analysis of our items as the second objective of the **study**.

In addition to considering DIF for focal and reference groups, we also look at the DIF between villages comprising the “control” and “intervention” groups which were pre-selected for the future phases of the study involving experiments. The control group consists of 7214 individuals living in 19 villages, and the intervention group consists of 7038 individuals living in 14 villages. If we find no DIF between these two groups, it will mean that the hope survey items are robust across different villages.

## 4. Context and Data

This study evaluates the baseline data that was collected in July-August 2020 by the World Vision (WV) team and its collaborators, Ifakara Health Institute and the University of Alberta. The large survey data collected is aimed at identifying the effectiveness of the EWV model and the types of improvements that EWV contributes to communities, as well as the paths to those changes; and to investigate the socio-cultural and spiritual communities, gender, political, and other contextual elements that facilitate or restrict the EWV's implementation and achievement of predicted best outcomes. EWV will be implemented in four World Vision Area Programmes (APs) namely Kihanga, Wasa, Kasanda, and Nyaronga, located in two regions in Tanzania.

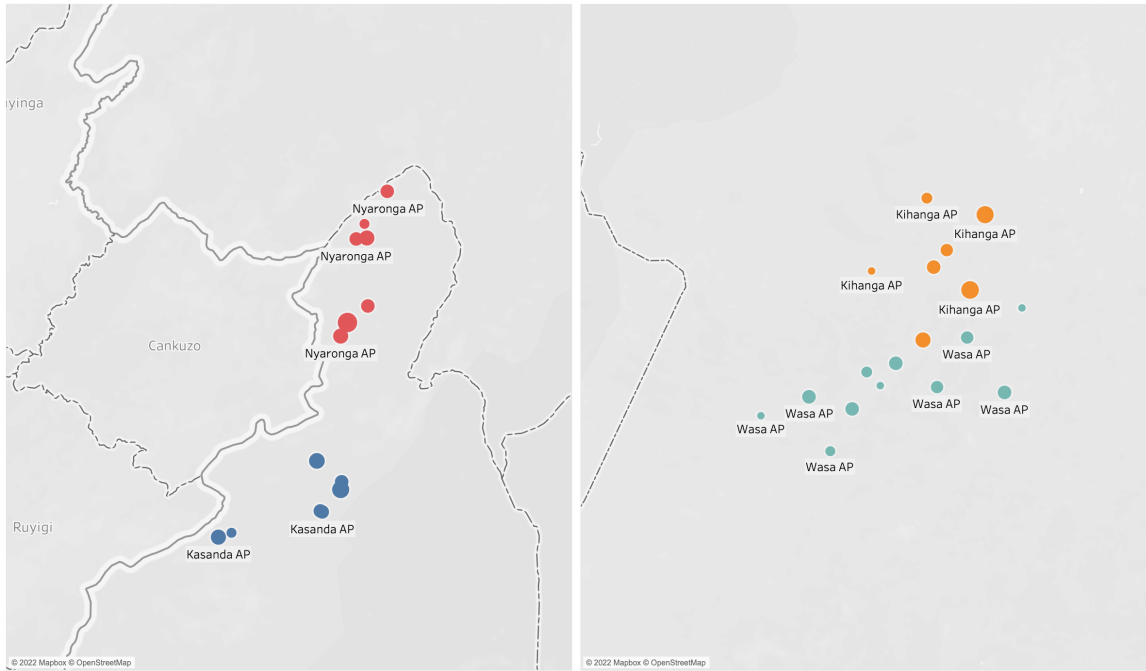


Figure 4.1: This map is a depiction of villages located in the study areas in Kigoma and Iringa. Kigoma: Kasanda AP (Control) and Nyaronga AP (Intervention; Iringa: Wasa AP (Control) and Kihanga AP (Intervention)

A two-stage cluster random sampling approach was used, with the first stage consisting of village selection and the second stage consisting of household sampling

within each village. Villages located at the border of the control and treatment areas were removed to reduce spill-over effects between treatment and control. A total of 98 villages were randomly selected, and a total of 2814 households were interviewed from those villages.

In Iringa and Kigoma, 1414 and 1400 households were interviewed. The average household size in Iringa is 4.6 people, whereas in Kigoma it is 5.3 people. 99% of households' religion in Iringa are Christian, while in Kigoma 88% are Christian and around 5% are Muslim. The breakdown of respondents by age and gender showed that in the category of adolescents (10-19 years old) males and females are almost equally represented, whereas in higher categories (20-49 years old) we have significantly more females than males. Half of the males are concentrated in the adolescent category. The females are more evenly spread out across age groups. In making inferences about males, we bear in mind that almost half of them are 19 or younger.

Gender	Female	3130 (59.9%)
	Male	2094 (40%)
Age (10-49 years)	Mean	26.09 (SD 11.76)
Region	Iringa	2756 (52.8%)
	Kigoma	2468 (47.2%)
Education	Total numbers of adolescents responded	491
	Never attended	20.60%
	Primary	55.78%
	Secondary	13.44%
Main Source of income	Sale/exchange of own produce (farming)	74.69%
	Wage employment (working for someone else)	11.33%
	Petty business/Vending/Trading	5.44%
	Labour (self-employed/e.g. driver)	3.24%
	Direct selling (e.g. cosmetics)	2.47%
	Other	3%
Total number of respondents		5,224

Table 4.1: Summary statistics concerning the socio-demographic characteristics of the respondents

Summary statistics of our sample in Table 4.1 indicate that 60% of respondents to the mindset questions are females, while 40% are males. The average age of respondents in our sample is 26. 20.6% of adolescents at the age 10-19 years have never attended a school. 74.7% of the households dedicate themselves to the sale/exchange of their own produced products from farming. 11.3% of a household's main income is from wage employment. Non-agricultural self-employment is relatively rare; around 11% of households have income from self-employed labor, direct selling, and petty businesses.

The Scioli's Hope Scale in Table 4.2 is a 12 item self-reported measure of hope, with 3 items in each measuring four components. In each component, there is a negative item like "worry that someone may betray me". Respondents are asked to rate the extent of agreement with these items on a 4-point Likert-type scale ranging from strongly disagree to strongly agree. Table 4.2 contains descriptive statistics for each of the hope items. The means for positive items fall into the categories of "agree" and "strongly agree" except for item 5. The means of the negative items fall in between "agree" and "disagree".

<b>Positive Items</b>					
	Mean	SD	Skew	Kurtosis	Correlation with total score
Note: 1-Strongly disagree; 2-Disagree; 3-Agree; 4-Strongly agree					
1. I feel loved by someone (attachment)	3.41	0.6	-0.72	0.79	0.55
2. There are people in my life that I completely trust (attachment)	3.31	0.65	-0.74	0.95	0.52
3. I'm making progress towards important goals (mastery)	3.27	0.59	-0.43	0.86	0.55
4. I have a purpose in life (mastery)	3.17	0.63	-0.42	0.52	0.58
5. I can handle any current or future difficulties (survival)	2.95	0.77	-0.74	0.63	0.34
6. The future will bring opportunities for a better life (survival)	3.04	0.65	-0.52	1.01	0.57
7. My faith in a higher power gives me the strength to pursue my dreams (spirituality)	3.25	0.64	-0.64	1.05	0.6
8. My spiritual beliefs have empowered me to succeed in life (spirituality)	3.3	0.64	-0.74	1.23	0.58
<b>Negative Items</b>					
Note: 1-Strongly agree; 2-Agree; 3-Disagree; 4-Strongly disagree					
9. I have never felt close to any kind of spiritual force or presence (spirituality)	2.85	0.86	-0.29	-0.64	0.34
10. I have doubts about achieving those things that really matter to me (mastery)	2.36	0.83	0.21	-0.48	0.37
11. I'm running out of options for improving my life (survival)	2.32	0.87	0.23	-0.6	0.32
12. I worry that someone may betray me (attachment)	2.57	0.92	-0.05	-0.83	0.36
N=5229					

Table 4.2: Item Mean, Standard Deviation (SD), Skewness, Kurtosis for hope items

Most of the items are skewed left, meaning that the participants tended to agree with the positive hope items and disagree with the negative hope items, which indicates high levels of hope overall. It is important to note that the negative items, items 9-12, are reverse coded, meaning that for these items a score of 1 means "Strongly agree" and a score of 4 means "Strongly disagree". By observation, the positive items have a higher mean than the negative items despite the reverse coding. This may

already suggest that the positive and negative items may address slightly different constructs. Furthermore, the last column of Table 4.2 shows the item's correlation with the total score, excluding that item from the survey, also known as item total correlation. It is a correlation between the individual item and the total score without that item. The total score is found by adding up the scores for each item. For example, for item 1 we sum up the score for the remaining 11 items for each person. As we see from the table, all the items have item total correlation scores above 0.3, meaning that all the items can be said to measure the same latent construct, i.e., hope. That being said, the negative items have lower values of item total correlation, indicating that they measure hope in a slightly different way. In other words, positively worded items and the corresponding negatively worded items may not be fully equivalent.

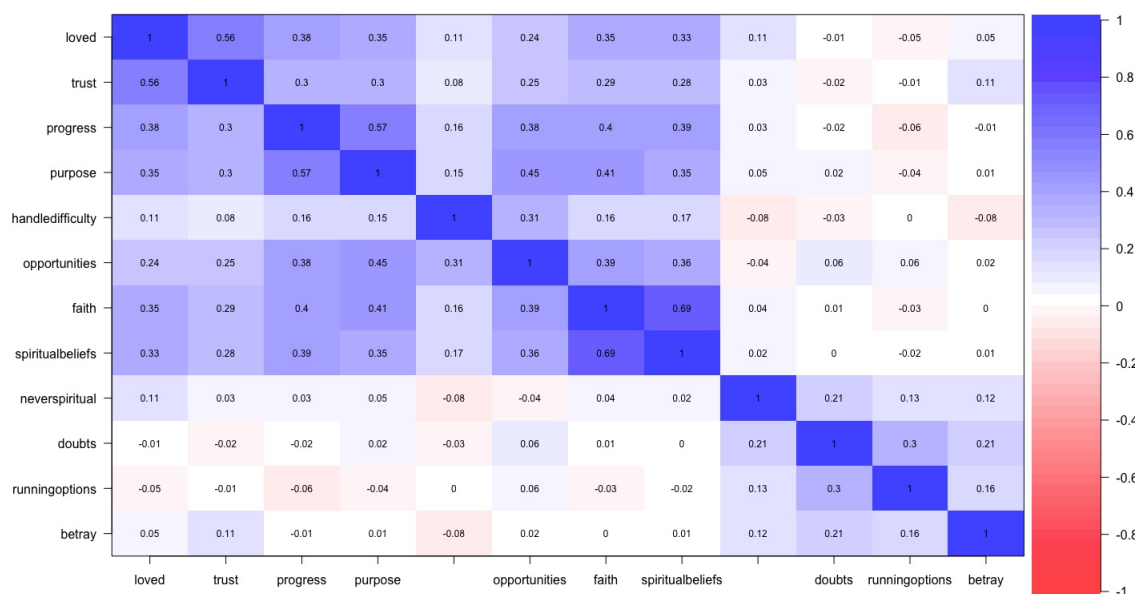


Figure 4.2: Pearson correlation for hope items

Figure 4.2 illustrates the Pearson correlation of the items graphically. Darker blue squares imply a higher correlation, whereas white and red-shaded squares indicate low or negative correlation, respectively. We see that the positive items are positively

correlated with each other. This supports the idea that the positive items represent the same latent construct. The negatively worded items, namely items 9-12, are reverse coded. Therefore, we should expect most of the squares to be blue-shaded. This would indicate that all the items represent the same construct. However, the boxes representing correlations between the positive and the negative items are mostly shaded white or red, again suggesting that the two groups are not perceived similarly by respondents. This is curious, because negative items being reverse coded, we would expect higher positive correlations between positive items and their respective negative counterparts. For example, item 8 is a positive spirituality item and item 9 is a negative spirituality item. If respondents strongly agreed with item 8, we would roughly expect them to strongly disagree with item 9 as well. With reverse coding this would have given us a high correlation between these items. On the contrary, we see almost a zero correlation between items 8 and 9. One possible explanation is that these items are not perceived as opposites by the respondents, and we should not treat them as positive-negative pairs.

Note also that item 5 relating to the ability to handle difficulty is least correlated with other positive items, except with item 6. In fact, we may notice that items relating to the same hope components are more strongly correlated with one another than with positive items from other hope components.

## 5. Results

We begin by examining the validity of the IRT assumptions. The first assumption is internal consistency which is required to make sure the hope scale contains items related only to the hope concept. Internal consistency measures whether several items that propose to measure the same construct produce similar scores. For example, if a respondent endorses a statement “I like riding bicycles” and disagrees with the statement “I hate bicycles”, this indicates a good internal consistency. Psychometric test validation techniques like Cronbach’s alpha are frequently used to measure internal consistency. Cronbach’s alpha determines the average correlation of its items.

Our estimates for Cronbach’s alpha values are presented in Table 5.1. Acceptable values of alpha are considered to range between 0.70 to 0.95 (Tavakol and Dennick 2011). From Table 5.1, our Cronbach’s alpha estimates for positive items fall in the range of acceptable levels of internal consistency. However, the alpha value estimate for the 12 items taken together falls under the range of acceptable values. This might be the result of a small number of questions or weak correlation. If a low alpha is caused by a lack of association between items, for example, some items should be changed or deleted. The simplest way is to determine the correlation between each test item and the overall score test; those with low correlations (close to zero) should be removed in order to only keep the items that only measure hope (Tavakol and Dennick 2011). Therefore, based on the results of Cronbach alpha, we

can only satisfy internal consistency for positive items taken separately, but not for the whole set of questions. Thus, we conduct the analysis for positive and negative items separately. There are examples in the literature of this being done, like Beaman et al. (Beaman 2012) who say that aggregating aspirational questions into one index may not be legitimate and may confound the effects of reservation on the various spheres of aspiration. Thus, after validation tests, they also excluded one poorly correlated question from the aspirations index (Beaman et al. 2012).

Another assumption of IRT is item invariance. If the invariance assumption holds true, items are understood the same way by all respondents. When this assumption does not hold, it might be because items were poorly worded or were interpreted differently by various samples. When an item's difficulty and discrimination, as defined in section 3.2, differ across subgroups even after controlling for the level of hope, it means that the invariance assumption does not hold and the item is said to exhibit differential item functioning (DIF). To recall, difficulty is the minimum latent trait level required to correctly answer an item and discrimination is an item's ability to distinguish between individuals with low and high latent trait levels. We will conduct a DIF analysis of our items as the second objective of the study.

One way to test the assumption of local independence is to look at the discrimination parameter (Hambleton, Swaminathan, and Rogers 1991; Edelen and Reeve 2007). If items exhibit excessive covariation or dependence, they may have extremely high slopes (e.g.  $>4$ ) in comparison to other items in the measure (Nguyen et al. 2014). To analyse possible violations of the assumption of local independence, the discrimination parameter was evaluated.

## Model choice and model fit

When designing our IRT study, we can assume that all of the items are equally effective at discriminating between respondents, or we can assume that each item has a unique discrimination parameter. Thus, we are comparing a “constrained” model to

	Cronbach's Alpha	95% CI
positive items	0.79	[0.78, 0.80]
negative items	0.48	[0.45, 0.50]
all items	0.64	[0.63, 0.66]

Table 5.1: Chronbach's Alpha

an “unconstrained” model - and, as we will see, the difference between these models in terms of their fit to the data can be explicitly tested. To compare the model fit statistics of constrained and non-constrained models, we use an ANOVA (Analysis of Variance) that evaluates the difference between the two models. We did it separately for positive and negative items, in Table 5.2.

<b>Fit of constrained vs non-constrained GRM for positive items of Hope with Likelihood ratio test</b>						
	AIC	BIC	log.Lik	LRT	df	p.value
Constrained model	67678.55	67842.6	-33814.27			
Non-constrained model	66697.88	66907.86	-33316.94	994.67	7	<0.001
<b>Fit of constrained vs non-constrained GRM for negative items of Hope with Likelihood ratio test</b>						
	AIC	BIC	log.Lik	LRT	df	p.value
Constrained model	52133.07	52218.75	-26053.53			
Non-constrained model	51980.36	52085.82	-25974.18	158.7	3	<0.001

Table 5.2: Fit on Constrained vs Non-constrained

A significant p-value tells us that the non-constrained model, when we do not assume that each item has the same discrimination parameter is a better fit to the data than the constrained model. Thus, we will use the GRM non-constrained model. Using the non-constrained model allows us to see which items discriminate respondents well by their hope levels, and which do not. The questions with a low discrimination ability, i.e. a flat slope, can then be removed as they poorly measure the hope level.

## 5.1 Objective 1: Item Response Theory for Hope

In the first objective we investigate whether the survey questions are measuring hope and not something else. To this aim we conduct GRM analysis and report results in Table 5.3 below. We separately consider positive and negative items and identify which questions discriminate respondents better according to their hope levels. Furthermore, having analyzed the location, or difficulty, of the item, we want to see if our items cover different parts of the hope scale, i.e., respondents with different hope levels are evenly represented. The parameters of discrimination and difficulty for each item of hope are presented in Table 5.3 below. Using the GRM to measure hope level, none of the discrimination factors exceeded 4.0. The discrimination parameters ( $a$ ) of the items range from 0.66 to 2.69 and the first ( $b1$ ), second ( $b2$ ) and third ( $b3$ ) difficulty parameters range from -4.38 to -1.78, -2.4 to 0.46, and 0.09 to 2.41, respectively. We use Baker and Kim's (2017) methodology for evaluating slope parameters, which hold under the logistic model for the item characteristic curve (Baker and Kim 2004, p.33). The discrimination parameter,  $a$ , equal to 0 means no ability for measuring latent trait,  $a$  between 0.01–0.04 means a very low ability,  $a$  between 0.35–0.64 means a poor ability,  $a$  between 0.65–1.34 means a moderate ability,  $a$  between 1.35–1.69 means a very high ability, and  $a$  greater than 1.70 signifies an extremely high level of ability (Baker and Kim 2004, p.33), where ability translates to hope in our case. All positive items except item five fall under high or very high discrimination, while three out of four negative items have moderate levels of discrimination. The lower discrimination parameters of negative items indicate that they distinguish individuals less accurately than the positive ones.

	a	b1	b2	b3
	(SE)	(SE)	(SE)	(SE)
<b>Positive statements</b>				
Item 1. I feel loved by someone.	1.74 (0.06)	-3.40 (0.12)	-2.40 (0.07)	0.09 (0.03)
Item 2. There are people in my life that I completely trust.	1.41 (0.05)	-3.53 (0.12)	-2.33 (0.07)	0.35 (0.03)
Item 3. I'm making progress towards important goals	2.17 (0.07)	-2.91 (-0.09)	-1.98 (0.05)	0.5 (0.02)
Item 4. I have a purpose in life.	2.05 (0.07)	-2.92 (-0.09)	-1.60 (0.04)	0.72 (0.03)
Item 5. I can handle any current or future difficulties.	0.76 (0.03)	-3.90 (-0.17)	-2.14 (0.10)	1.93 (0.10)
Item 6. The future will bring opportunities for a better life.	1.73 (0.05)	-2.87 (-0.08)	-1.51 (0.04)	1.14 (0.03)
Item 7. My faith in a higher power gives me the strength to pursue my dreams.	2.69 (0.09)	-2.44 (-0.06)	-1.66 (0.04)	0.44 (0.02)
Item 8. My spiritual beliefs have empowered me to succeed in life.	2.43 (0.08)	-2.53 (-0.07)	-1.82 (0.04)	0.33 (0.02)
<b>Negative Statements</b>				
Item 9. I have never felt close to any kind of spiritual force or presence.	0.66 (0.04)	-4.38 (0.36)	-1.15 (0.08)	1.9 (0.12)
Item 10. I have doubts about achieving those things that really matter to me.	1.80 (0.11)	-1.54 (0.06)	0.30 (0.03)	1.85 (0.07)
Item 11. I'm running out of options for improving my life.	1.1 (0.06)	-1.78 (0.07)	0.46 (0.04)	2.41 (0.1)
Item 12. I worry that someone may betray me.	0.78 (0.04)	-2.70 (0.14)	-0.18 (0.04)	2.31 (0.12)

Table 5.3: IRT, Graded Response Model results for Hope

In Figure 5.1 we graph the results of the Graded Response Model for each of the 12 items that we had in Table 5.3. The vertical axis is the probability of the positive response to the item, or the agreement with the item. The horizontal axis is the level

of hope: the higher the value of the x-axis, the higher the hope level. Each graph consists of four curves, each corresponding to the level of agreement with the item, namely strongly disagree, disagree, agree, and strongly agree. Each curve shows the probability of agreeing with the item given the level of hope. For example, the solid line graph corresponds to “strongly agree”. We see in 5.1(a), which is a graph for item 1, that people with higher levels of hope were more likely to agree with item 1 than people with low levels of hope. Ideally we want each of the curves to form four distinct bumps. This way, each level of agreement with the item would correspond to its own level of hope. We can see this in graphs for items 1, 3, 4, 6, 7, 8, and 10. In contrast, items 5, 9, 11 and 12 have overlapping curves, which makes it hard to map levels of agreement with the item to the hope levels. For example, a person with hope level -1 in the graph for item 9 is equally likely to agree or disagree and strongly agree or strongly disagree with the item.

Further, we can compare positive items (items 1-8) and negative items (items 9-12). We can see that all positive items except item 5 discriminate quite well. For item 5, we see many respondents “agree” with the item regardless of their hope level. For negative items, we see that in general the answer categories do a worse job of discriminating respondents by hope level. For example in items 9 and 12, given a level of hope, there are similar chances of agreeing or disagreeing with the item.

Furthermore, respondents need higher levels of hope to disagree with negative statements than to agree with positive statements. For example, Item 8 requires a low level of hope to elicit strong agreement, but Item 11 would seem to require a much higher level of hope to be endorsed. That is, even people who are only somewhat hopeful (i.e., people who have relatively low levels of hope) would likely agree with item 8, “My spiritual beliefs have empowered them to succeed in life”. In contrast, a person would probably need to be very hopeful to disagree with the item 11 “I am running out of options for improving their life”. Looking at the slopes/discrimination parameters in Table 5.3 and Figure 5.1, we can say that disagreeing with negative

statements requires more effort towards increasing hope levels, compared to agreeing with positive statements. For example, on a math test, you may need less effort to increase your ability to correctly solve easy questions, while hard questions require more time and effort to reach the ability level at which you can solve them. In our case, for example, disagreeing with Item 11, “I am running out of options for improving my life”, requires a substantial upgrade in hope levels compared to agreeing with Item 4 “I have a purpose in life.” We also see from Figure 5.1 that in Items 5, 9 and 12 the distributions of the categories significantly overlap, meaning that individuals with very different levels of hope have a similar probability of choosing a specific category. In effect, these items do a poor job of discriminating between high and low hope individuals. Interestingly, the difficulty of negative items shown in Table 5.3 goes up to 2.41, indicating a variation among individuals with higher levels of hope, which adds substantially to positive ones. In other words, negative items were the “hardest” items for the respondents to endorse. This suggests that negative items target higher trait levels of hope than positive items.

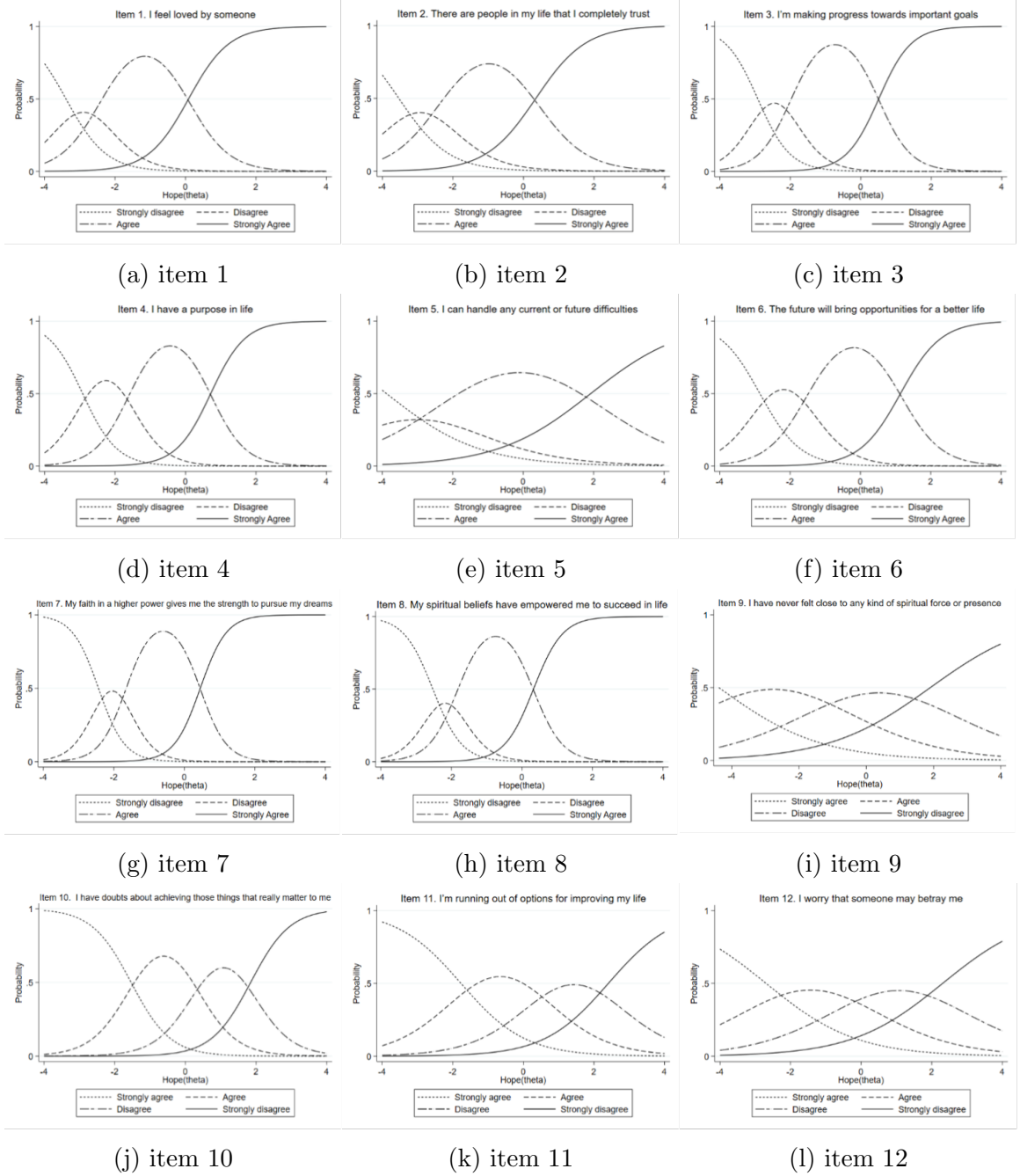


Figure 5.1: IRT for Hope

### 5.1.1 Item Information Functions and Test Information Functions for Hope

In this section we consider the informativeness of the items across different levels of hope. Figures 5.2 (a) and (b) present item information curves for each positive and

negative item. The height of the curve indicates the amount of information that the item provides. The highest point on a curve represents the hope level at which the item provides the most information. In fact, an item provides the most information at a trait level that corresponds with its difficulty level, as estimated earlier. For example, Item 7, “My faith in a higher power gives me the strength to pursue my dreams”, is one of the easiest items and provides the best information at a trait level of -2. In general, the figures show that positive items are good at measuring latent hope since they provide more information, but only at lower and average levels of hope. In other words, positive items are more capable of discriminating among individuals with lower hope levels than among people with high hope levels, presumably because most people with high hope levels agree with the statements. Figures 5.2 (c) and (d) show the graphs for the test information functions and their corresponding standard errors (SE) for the hope scale. As can be seen in this graph, the SE values were relatively low for hope values ranging from -4 to 1.5 for positive items and from -2 to 2 for negative items. Thus, positive items provide strong and precise information on hope except at high levels of hope. As for the negative items, they provide weak information which is imprecise at low and high levels of hope. Recalling our first objective which is to check whether our survey is measuring what we intend to measure, we also want to see whether it measures different levels of hope. Looking at Figure 5.2 (a) and (b), note that items 5, 9, 11 and 12 provide very little information. It may be a good idea to replace them with items that provide more information, especially about the average and high levels of hope ( $\theta$  between 1 and 3). We see that, in general, negative items are less informative, except for Item 10, which provides good information and covers people with a wider range of hope levels, especially those who have higher levels of hope. Similar findings were reported by Zanon et al. (2016), who discovered that surveys with positive and negative questions may be differently perceived. They found that in general positive statements explain latent constructs better than negative statements (Zanon et al. 2016). Positive statements are more

informative at lower levels of the latent trait, while negative statements can still be useful at higher levels.

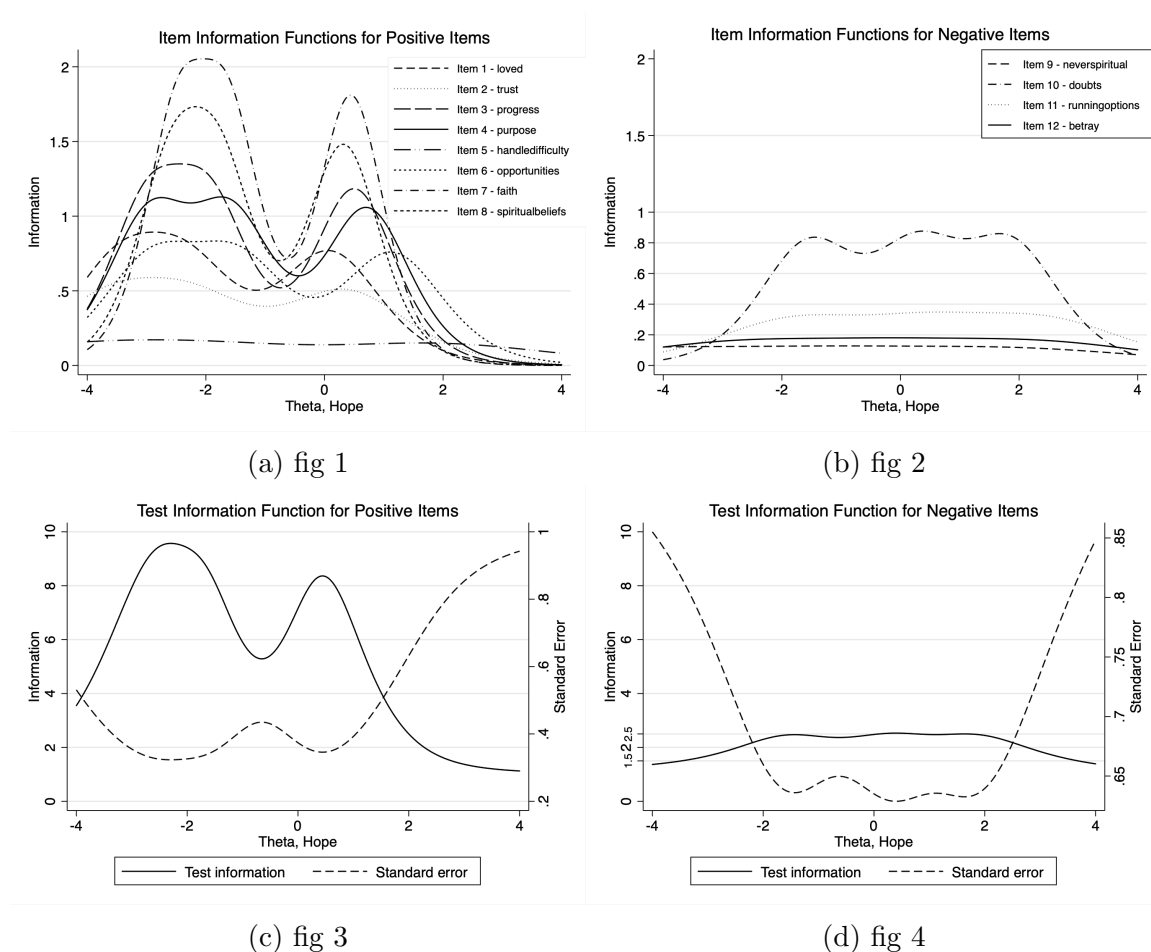


Figure 5.2: Item Information Function (IIF) and Test Information Function (TIF) of the Scioli Hope. Latent trait is shown on the horizontal axis, and the amount of information for each item for IIF and the amount of information for all items and standard error yielded by the test at any trait level are shown on the vertical axis

If we use a shortened survey with positive items, it will be better at assessing individuals with low to moderate levels of hope, but not as good in assessing individuals with higher levels of hope.

## 5.2 Objective 2: Differential Item Functioning Results for Hope

To examine whether items behaved equivalently across a range of different subgroups of people, we conduct differential item functioning (DIF) analyses, which also comprise a test of the invariance assumption of the IRT. In DIF analyses, six grouping criteria were examined: gender, region, age, whether respondents had access to food and water, and treatment. We divided them into two groups, focal and reference, as explained in Section 3.4: the “focal” group is the one suspected to be at a disadvantage to the “reference” group. We divide the sample into reference and focal groups, respectively, like so: males vs females, adults vs adolescents (10-19 years old), Kigoma vs Iringa; those who for the past one year met their daily food needs and those who didn’t; those who have access to protected water or have their own tap water at home versus those who use unprotected sources of water. We also consider the DIF for the control and intervention groups, as defined in Section 3.4. DIF analyses compare item endorsement rates in the focal group compared to the reference group, conditional on the latent trait. For example, we would compare respondents who have no access to protected water (focal group) compared to the rest who have access to any type of protected water (reference group) given the same level of the trait. When respondents with the same overall hope score have differing probabilities of endorsing an individual item, the item is said to demonstrate DIF; in other words, the item acts differentially between the two groups.

	Gender (SE) (1-male; 2- female)	Region (0-Kigoma; 1-Iringa)	Age (SE) (0-adoles (10- 19 years); 1- adult)	Food Security (0- secure, 1- no)	Access to protected water (0=yes, 1=no)	Treatment (0-control; 1- intervention)
Item 1. I feel loved by someone.	-0.04 (0.07)	-0.62 (0.08) ***	-0.24 (0.07) **	0.02 (0.09)	0.12 (0.08)	0.05 (0.03)
Item 2. There are people in my life that I completely trust.	-0.1 (0.06)	-0.64 (0.07) ***	-0.51 (0.07) ***	-0.14 (0.08) *	0.06 (0.07)	0.07 (0.03) **
Item 3. I'm making progress towards important goals	-0.18 (0.08) **	0.09 (0.08)	-0.04 (0.08)	0.08 (0.1)	-0.20 (0.08) **	-0.04 (0.04)
Item 4. I have a purpose in life.	-0.09 (0.07)	0.58 (0.08) ***	0.19 (0.07) **	0.22 (0.09) **	0.15 (0.08) **	-0.04 (0.25)
Item 5. I can handle any current or future difficulties.	-0.25 (0.06) ***	-0.15 (0.06) **	1.14 (0.06) ***	-0.26 (0.07) ***	-0.03 (0.06)	-0.03 (0.03)
Item 6. The future will bring opportunities for a better life.	-0.17 (0.07) **	0.8 (0.07) ***	0.08 (0.07)	0.13 (0.09)	0.01 (0.07)	-0.01 (0.03)
Item 7. My faith in a higher power gives me the strength to pursue my dreams.	0.15 (0.08) *	0.13 (0.12)	1.14 (0.08) *	-0.3 (0.1)	0.18 (0.09) **	0.02 (0.04)
Item 8. My spiritual beliefs have empowered me to succeed in life.	0.25 (0.08) **	-0.11 (0.08)	-0.15 (0.08)	-0.32 (0.1) **	-0.01 (0.08)	-0.03 (0.04)
Item 9. I have never felt close to any kind of spiritual force or presence.	0.07 (0.05)	0.09 (0.06) *	0.09 (0.05) *	0.09 (0.07)	-0.04 (0.06)	-0.04 (0.03)
Item 10. I have doubts about achieving those things that really matter to me.	0.01 (0.05)	0.27 (0.06) ***	0.16 (0.05) **	-0.45 (0.07) ***	0.04 (0.06)	-0.00 (0.03)
Item 11. I'm running out of options for improving my life.	0.03 (0.05)	0.42 (0.06) ***	-0.37 (0.05) ***	-0.38 (0.07) ***	-0.14 (0.06) **	0.04 (0.03) *
Item 12. I worry that someone may betray me.	0.04 (0.05)	-0.15 (0.06) **	-0.27 (0.05) ***	-0.17 (0.65) **	0.02 (0.06)	-0.02 (0.03)

Note: \*= $p<0.1$ ; \*\*= $p<0.05$ ; \*\*\*= $p<0.01$

Food security: In the past 12 months how did your household meet daily food needs?

Water access: What is the main water source for your household?

Table 5.4: DIF for gender, age, region, food security, access to water and treatment for Hope

The absence of DIF indicates that the probability of a person's response to these items depends only on her or his hope level and not on other factors. In the context of the measurement of hope for females and males, a person's response to an item will depend only on the level of hope she or he has and not on whether the person is female or male.

Table 5.4 presents DIF analyses of the survey items conditional on hope. From the

table we see that there is no DIF for items 3, 4, 6 and 7 among adults and adolescents, for items 3, 7, 8, 9 and 12 among people living in Kigoma and Iringa, for items 3, 4, 7 and 9 among people who could meet their daily food needs and those who could not, for items 4, 5, 7, 8, 9, 10 and 12 among those who have access to protected water or have tap water at home and those who access unprotected sources of water, and for 10 items for the treatment group.

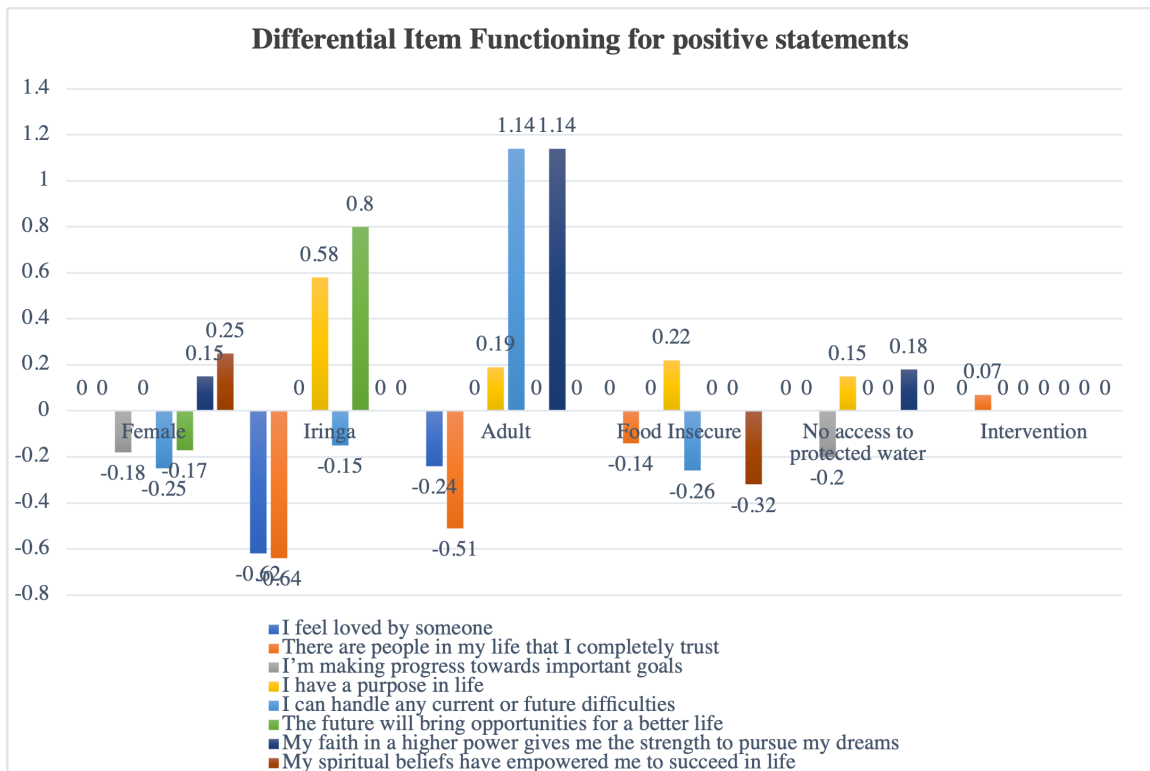


Figure 5.3: DIF for positive items with significant DIF

At the same time the results demonstrate some DIF for items by groups. To make the results visually clearer we present Figure 5.3 which includes only positive items with statistically significant DIF. The same is done for negative items in Figure 5.4. From the figure we see that females, given the same hope level as males, are less likely to agree that they are making progress, that they can handle current and future difficulties, and that the future will bring opportunities for a better life. However, they have a stronger belief in faith and spirituality that will bring them closer to

their dreams and success compared to males with the same hope levels. Adults are more likely to agree that they can handle difficulties and that their faith in a higher power gives them the strength to pursue their dreams. However, adolescents with the same level of hope as adults are more likely to agree that they feel loved and have people whom they can trust. Moreover, DIF analyses suggest that people who do not have access to protected water are less likely to agree with the statement “I’m making progress towards important goals.” Those who managed to meet their daily food needs for the last year and have access to water are also more likely to say that they have a purpose in life, even though they have equal hope levels with those who are food insecure and don’t have access to protected water. There is also DIF by region, as seen in Figure 5.3. People living in Iringa, given the same level of hope as people from Kigoma, are less likely to agree that they feel loved by someone and have people whom they can completely trust. However, in Iringa, people are more likely to have a purpose in life and believe that the future will bring opportunities for a better life.

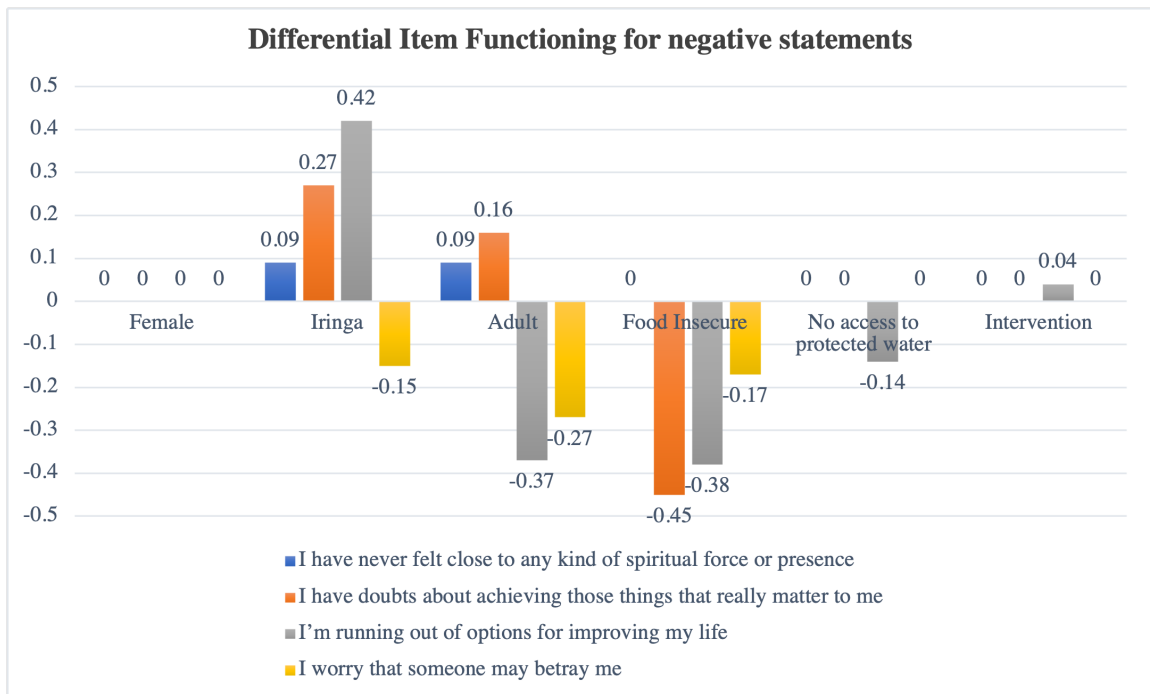


Figure 5.4: DIF for negative items with significant DIF

Figure 5.4 visually presents negatively stated items with significant DIF. Results suggest that people living in Iringa are more likely to disagree that they have never felt close to any kind of spiritual force or have doubts about achieving goals or running out of options for improving life, while they worry more that someone may betray them. Also, adults are less likely to disagree that they are running out of options and that someone may betray them. Those who didn't manage to meet their daily food needs for the last year were also less likely to disagree with three negative statements.

## 6. Conclusions

Overall, IRT analyses demonstrate that the items on the hope measure show strong relationships with the hope as they adequately discriminate between individuals with different levels of hope. In the Tanzanian context, survey items are most informative for individuals with low and average levels of hope and least informative at higher levels of hope. It means that we can differentiate people at low and average levels of hope but not yet at higher hope levels. Thus, we suggest that future research include more items that are informative at higher hope levels.

The analysis of differential item functioning (DIF) for hope reveals differences across groups. We see almost no DIF for the treatment and control groups in the measurement of hope, which means that the groups are comparable enough for conducting RCTs in the future phases of the study. However, there is some DIF by region and between adults and adolescents. People living in Iringa are less likely to agree that they feel loved and have people whom they can trust but are more likely to have a purpose in life and believe in future opportunities. Adults with the same level of hope as adolescents are more likely to agree that they can handle difficulties and that their faith in a higher power gives them the strength to pursue their dreams. However, adolescents are more likely to agree that they feel loved and have people whom they can trust.

Empirical economists emphasize the importance of comparing apples to apples, i.e.

holding all else constant as the effect of an intervention is assessed. The increasing use of randomized experiments ushered in by the credibility revolution in economics brought us closer to making such adequate comparisons. However, as economists venture into studying the effects of latent constructs like hope, which were traditionally the domain of other social sciences, they inadvertently make apples-to-oranges comparisons by not using proper measurement tools. On the surface, the word “hope” has the same meaning to different people, but in reality its understanding varies with culture, geography, upbringing and many other factors. While psychologists and sociologists are wary of these discrepancies and use statistical techniques to keep their measurement tools in check, economists are just beginning to appreciate the complexity of measuring latent constructs.

This study analyzed baseline survey data collected in Tanzania with the aim of appraising the robustness of our survey items to measurement errors in order to make necessary corrections for the further stages of the study involving randomized experiments. The hope survey was adapted from an existing hope scale typically used in developed countries, and thus its ability to measure hope needed to be tested in the context of Tanzania first before proceeding with the later stages of the study. The first objective of the study was to identify if survey items can successfully distinguish respondents by hope levels. The second objective was to see whether survey items were interpreted the same way or differently across different population categories.

The first objective was addressed in Section 5.1. We find that our survey items indeed can successfully discriminate survey participants by their levels of hope. Using the methodology outlined in Baker and Kim (2004) we find that all survey items have at least a moderate ability to discriminate respondents by hope level. At the same time, positive items discriminate better than negative items. In fact, all positive items except item 5 demonstrated a high or very high discrimination ability, while three out of four negative items have moderate levels of discrimination, and the fourth having a very high discrimination ability. However, we find that in the Tanzanian context,

survey items are most informative for individuals with low and average levels of hope and least informative at higher levels of hope. It means that we can differentiate people at low and average levels of hope but not yet at higher hope levels. Thus, we suggest that future research include more items that are informative at higher hope levels.

The second objective was addressed in **Section 5.2**. Overall, we find that there are differences in how groups of respondents answered the survey. First of all, we see almost no DIF for the treatment and control groups in the measurement of hope, which means that the groups are comparable enough for conducting RCTs in the future phases of the study. However, we find high DIF in the responses of adults and adolescents. If this DIF is ignored in the future stages of the study, adolescents will be erroneously scored as less hopeful than adults. This is because of two items in our study with which adults would likely agree with if they were hopeful, but agreeing with these items would be indicative of a high hope level for adolescents. One way to address this problem is to remove the items contributing with high DIF. This is not necessarily the best procedure because item removal can change the meaning of the construct and lowered reliability (Teresi 2006; Hambleton 2006), i.e. the remaining items may not be enough to encapsulate hope. Moreover, some of our items show DIF cancellation, a situation when DIFs cancel each other out, and their removal could bias the results. This happens when we compare Iringa and Kigoma for most of the items. An alternative solution is to compare results within groups and not across groups.

# Bibliography

- Adams V.H., Snyder C.R. Rand K.L. King E.A. Sigman D.R. and K.M. Pulvers (2002). *Hope in the Workplace*. In Giacalone, R. and Jurkiewicz, C., Eds., *Handbook of Workplace Spirituality and Organization Performance*, Sharpe, New York.
- Anderson, Craig L. and David B. Feldman (2020). “Hope and Physical Exercise: The Contributions of Hope, Self-Efficacy, and Optimism in Accounting for Variance in Exercise Frequency”. In: *Psychological reports* 123.4, pp. 1145–1159. ISSN: 1558-691X. DOI: 10.1177/0033294119851798. URL: <https://pubmed.ncbi.nlm.nih.gov/31142190/>.
- Appadurai, Arjun (2004). “Culture and Public Action. In Rao, Vijayendra and Michael Walton (ed). *The Capacity to Aspire: Culture and the Terms of Recognition*”. In: *Culture*, pp. 59–84.
- Aspinwall, Lisa G. and Samantha L. Leaf (2002). “In search of the unique aspects of hope: Pinning our hopes on positive emotions, future-oriented thinking, hard times, and other people”. In: *Psychological Inquiry* 13.4, pp. 276–288. ISSN: 1047840X. DOI: 10.1207/S15327965PLI1304.
- Baker, Frank B and Seock-Ho Kim (2004). *Item response theory: Parameter estimation techniques*. CRC press.
- Bandura, Albert (1982). “Self-efficacy mechanism in human agency”. In: *American Psychologist* 37.2, pp. 122–147. ISSN: 0003066X. DOI: 10.1037/0003-066X.37.2.122.
- Bandura, Albert et al. (2019). “Leading an African Renaissance”. In: *Leading an African Renaissance* 84.2, pp. 133–154. ISSN: 00219010. DOI: 10.1037/1099-9809.13.2.94. URL: <https://doi.org/10.1080/13674676.2019.1594740>[https://via.library.depaul.edu/csh\\_etd/296http://dx.doi.org/10.1016/j.clinthera.2014.04.006](https://via.library.depaul.edu/csh_etd/296http://dx.doi.org/10.1016/j.clinthera.2014.04.006).
- Banerjee, Abhijit et al. (2015). “A multifaceted program causes lasting progress for the very poor: Evidence from six countries”. In: *Science* 348.6236. ISSN: 10959203. DOI: 10.1126/science.1260799.
- Beaman, Lori et al. (2012). “Female leadership raises aspirations and educational attainment for girls: A policy experiment in India”. In: *Science* 335.6068, pp. 582–586. ISSN: 10959203. DOI: 10.1126/science.1212382.
- Bénabou, Roland et al. (n.d.). “Self-knowledge and self-regulation: An economic approach”. In: *Citeseer* (). URL: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.448.5005&rep=rep1&type=pdf>.
- Berg, Carla J. et al. (2011). “The role of hope in engaging in healthy behaviors among college students”. In: *American Journal of Health Behavior* 35.4, pp. 402–415. ISSN: 10873244. DOI: 10.5993/AJHB.35.4.3.

- Besley, Timothy (2017). "Aspirations and the political economy of inequality". In: *Oxford Economic Papers* 69.1, pp. 1–35. ISSN: 14643812. DOI: 10.1093/oep/gpw055.
- Birnbaum, A. (1968). "Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. Statistical Theories of Mental Test Scores". In.
- Bloem, Jeffrey and Bruce Wydick (2021). "All I Really Need to Know I Learned In Kindergarten? Evidence from the Philippines". In: *Economic Development and Cultural Change*. ISSN: 0013-0079. DOI: 10.1086/715502.
- Bloem, Jeffrey R. et al. (2018). "Measuring Hope: A Quantitative Approach with Validation in Rural Myanmar". In: *Journal of Development Studies* 54.11, pp. 2078–2094. ISSN: 17439140. DOI: 10.1080/00220388.2017.1385764. URL: <https://doi.org/10.1080/00220388.2017.1385764>.
- Chen, Fang Fang et al. (2012). "Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches". In: *Journal of Personality* 80.1, pp. 219–251. ISSN: 00223506. DOI: 10.1111/j.1467-6494.2011.00739.x.
- Chen, Gilad, Stanley M. Gully, and Dov Eden (2001). "Validation of a New General Self-Efficacy Scale". In: <https://doi.org/10.1177/1094428101410041>, pp. 62–83. ISSN: 10944281. DOI: 10.1177/1094428101410041.
- Core Guide: Measure Development*, (2020). Tech. rep. URL: <https://sites.globalhealth.duke.edu/rdac/wp-content/uploads/sites/27/2020/08/Core-Guide-Measure-Design-05-04-20.pdf>.
- Darolia, Rajeev and Bruce Wydick (2011). "The Economics of Parenting, Self-esteem and Academic Performance Theory and a Test". In: *Economica* 78.310, pp. 215–239. ISSN: 00130427. DOI: 10.1111/j.1468-0335.2009.00802.x.
- De Sá, Antonio Reis et al. (2019). "Can gender and age impact on response pattern of depressive symptoms among college students? A differential item functioning analysis". In: *Frontiers in Psychiatry* 10.FEB. ISSN: 16640640. DOI: 10.3389/fpsy.2019.00050.
- Dorans, Neil J. and Neal M. Kingston (1985). "the Effects of Violations of Unidimensionality on the Estimation of Item and Ability Parameters and on Item Response Theory Equating of the Gre Verbal Scale". In: *Journal of Educational Measurement* 22.4, pp. 249–262. ISSN: 17453984. DOI: 10.1111/j.1745-3984.1985.tb01062.x.
- Drasgow, Fritz and Charles K. Parsons (1983). "Application of Unidimensional Item Response Theory Models to Multidimensional Data". In: *Applied Psychological Measurement* 7.2, pp. 189–199. ISSN: 15523497. DOI: 10.1177/014662168300700207.
- Du, Hongfei and Ronnel B. King (2013). "Placing hope in self and others: Exploring the relationships among self-construals, locus of hope, and adjustment". In: *Personality and Individual Differences* 54.3, pp. 332–337. ISSN: 01918869. DOI: 10.1016/J.PAID.2012.09.015.
- Dufo, Esther (2012). *Lack of Hope and Persistence of Poverty*. URL: [https://www.econ.cam.ac.uk/events-files/seminars/MarshallLecture/2012-13/hope\\_cambridge.pdf](https://www.econ.cam.ac.uk/events-files/seminars/MarshallLecture/2012-13/hope_cambridge.pdf).

- Duggal, Devika, Amanda Sacks-Zimmerman, and Taylor Liberta (2016). "The Impact of Hope and Resilience on Multiple Factors in Neurosurgical Patients". In: *Cureus* 8.10. ISSN: 2168-8184. DOI: 10.7759/cureus.849.
- Dunn, Thomas J., Thom Baguley, and Vivienne Brunsden (2014). "From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation". In: *British Journal of Psychology* 105.3, pp. 399–412. ISSN: 20448295. DOI: 10.1111/BJOP.12046.
- Edelen, Maria Orlando and Bryce B. Reeve (2007). "Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement". In: *Quality of Life Research* 16.SUPPL. 1, pp. 5–18. ISSN: 09629343. DOI: 10.1007/S11136-007-9198-0.
- Embretson S. E., Reise S. P. (2000). "Item response theory for psychologists". In: URL: <https://psycnet.apa.org/record/2000-03918-000>.
- Faith and Development in Focus: Tanzania*, (2018). Tech. rep. URL: <https://berkeleycenter.georgetown.edu/publications/faith-and-development-in-focus-tanzania>.
- Franco, Debra L et al. (2008). "Motivation, self-efficacy, physical activity and nutrition in college students: Randomized controlled trial of an internet-based education program". In: *Preventive medicine* 47.4, pp. 369–377.
- Furr, R. Michael (2014). "Scale Construction and Psychometrics for Social and Personality Psychology". In: *Scale Construction and Psychometrics for Social and Personality Psychology*. DOI: 10.4135/9781446287866.
- Gallagher, Matthew W. and Shane J. Lopez (2009). "Positive expectancies and mental health: Identifying the unique contributions of hope and optimism". In: *Journal of Positive Psychology* 4.6, pp. 548–556. ISSN: 17439760. DOI: 10.1080/17439760903157166.
- Genç, Emel and Gökmen Arslan (2021). "Optimism and dispositional hope to promote college students' subjective well-being in the context of the COVID-19 pandemic". In: *Journal of Positive School Psychology* 5.2, pp. 87–96. ISSN: 27177564. DOI: 10.47602/JPSP.V5I2.255.
- Glewwe, Paul, Phillip H Ross, and Bruce Wydick (2014). "Developing Aspirations: The Impact of Child Sponsorship on Self-Esteem and Life Expectations". In: *University of San Francisco*, pp. 1–34.
- Hambleton, RK, H Swaminathan, and HJ Rogers (1991). *Fundamentals of item response theory*.
- Hambleton, Ronald K (2006). "Good practices for identifying differential item functioning". In: *Medical care* 44.11, S182–S188.
- Hutz, Claudio S et al. (2014). "The Relationship of Hope, Optimism, Personality in Brazilians and Americans". In: *Psychology* 5.April, pp. 514–522.
- IRIS, (2017). Tech. rep. URL: <https://www.iris-france.org/wp-content/uploads/2017/03/Tanzania-Final-report.pdf>.
- Jensen, Robert and Emily Oster (2009). "The power of TV: Cable television and women's status in India". In: *Quarterly Journal of Economics* 124.3, pp. 1057–1094. ISSN: 00335533. DOI: 10.1162/qjec.2009.124.3.1057.

- Jerusalem M., Schwarzer R. (1992). *Self-efficacy as a resource factor in stress appraisal processes*. URL: <https://psycnet.apa.org/record/1992-97719-009>.
- Kleist, Nauja and Stef Jansen (2016). "Introduction: Hope over Time—Crisis, Immobility and Future-Making". In: *History and Anthropology* 27.4, pp. 373–392. ISSN: 02757206. DOI: 10.1080/02757206.2016.1207636.
- Luszczynska, Aleksandra, Urte Scholz, and Ralf Schwarzer (2005). "The general self-efficacy scale: Multicultural validation studies". In: *Journal of Psychology: Interdisciplinary and Applied* 139.5, pp. 439–457. ISSN: 19401019. DOI: 10.3200/JRLP.139.5.439-457.
- Luthans, Fred and Susan M. Jensen (2002). "Hope: A New Positive Strength for Human Resource Development". In: *Human Resource Development Review* 1.3, pp. 304–322. ISSN: 15344843. DOI: 10.1177/1534484302013003.
- Lybbert, Travis J. and Bruce Wydick (2018). "Poverty, aspirations, and the economics of hope". In: *Economic Development and Cultural Change* 66.4, pp. 709–753. ISSN: 00130079. DOI: 10.1086/696968.
- Lybbert, Travis J et al. (2016). "Poverty, Aspirations, and the Economics of Hope: A Framework for Study with Preliminary Results from the Oaxaca Hope Project". In.
- Marlon, Jennifer R. et al. (2019). "How Hope and Doubt Affect Climate Change Mobilization". In: *Frontiers in Communication* 4.May, pp. 1–15. ISSN: 2297900X. DOI: 10.3389/fcomm.2019.00020.
- National Bureau of Statistics (2014). Tech. rep. URL: [https://www.tanzania.go.tz/egov\\_uploads/documents/TANZANIA\\_MAINLAND\\_SOCIO\\_ECONOMIC\\_PROFILE.sw.pdf](https://www.tanzania.go.tz/egov_uploads/documents/TANZANIA_MAINLAND_SOCIO_ECONOMIC_PROFILE.sw.pdf).
- Neumark-Sztainer, Dianne et al. (2003). "Correlates of fruit and vegetable intake among adolescents: Findings from Project EAT". In: *Preventive medicine* 37.3, pp. 198–208.
- Nguyen, Tam H. et al. (2014). *An introduction to item response theory for patient-reported outcome measurement*. DOI: 10.1007/s40271-013-0041-0. URL: <https://pmc/articles/PMC4520411/> <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4520411/?report=abstract> <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4520411/>.
- Nothwehr, F., D. O. Clark, and A. Perkins (2013). "Hope and the Use of Behavioral Strategies Related to Diet and Physical Activity". In: *Journal of human nutrition and dietetics : the official journal of the British Dietetic Association* 26.0 1, p. 159. ISSN: 09523871. DOI: 10.1111/JHN.12057.
- Ray, Debraj (2006). "Aspirations, poverty, and economic change". In: *Understanding Poverty* March. DOI: 10.1093/0195305191.003.0028.
- Redlich-Amirav, Dorit et al. (2018). "Psychometric properties of Hope Scales: A systematic review". In: *International journal of clinical practice* 72.7. ISSN: 1742-1241. DOI: 10.1111/IJCP.13213. URL: <https://pubmed.ncbi.nlm.nih.gov/29920883/>.
- Reese, L. M. (1999). "A Classical Test Theory Perspective on LSAT Local Item Dependence. LSAC Research Report Series. Statistical Report". In.
- Reise, Steven P. et al. (2013). "Multidimensionality and Structural Coefficient Bias in Structural Equation Modeling: A Bifactor Perspective". In: *Educational*

- and *Psychological Measurement* 73.1, pp. 5–26. ISSN: 15523888. DOI: 10.1177/0013164412449831.
- Rustøen, Tone, Bruce A. Cooper, and Christine Miaskowski (2011). “A longitudinal study of the effects of a hope intervention on levels of hope and psychological distress in a community-based sample of oncology patients”. In: *European Journal of Oncology Nursing* 15.4, pp. 351–357. ISSN: 14623889. DOI: 10.1016/j.ejon.2010.09.001. URL: <http://dx.doi.org/10.1016/j.ejon.2010.09.001>.
- Samejima, Fumiko (1969). “Estimation of latent ability using a response pattern of graded scores”. In: *Psychometrika* 34.S1, pp. 1–97. ISSN: 0033-3123. DOI: 10.1007/BF03372160.
- Schweizer, Karl (2011). “On the changing role of Cronbach’s  $\alpha$  in the evaluation of the quality of a measure”. In: *European Journal of Psychological Assessment* 27.3, pp. 143–144. ISSN: 10155759. DOI: 10.1027/1015-5759/A000069.
- Scioli, A et al. (2011). “Hope: Its nature and measurement”. In: *Psychology of Religion and Spirituality* 3.2, pp. 78–97. ISSN: 19411022. DOI: 10.1037/a0020903.
- Sen, Amartya (2014). “Development as freedom (1999)”. In: *The globalization and development reader: Perspectives on development and global change* 525.
- Sherer, Mark et al. (1982). “The Self-Efficacy Scale: Construction and Validation.” in: <http://dx.doi.org/10.2466/pr0.1982.51.2.663> 51.2, pp. 663–671. ISSN: 0033-2941. DOI: 10.2466/PR0.1982.51.2.663. URL: <https://journals.sagepub.com/doi/10.2466/pr0.1982.51.2.663>.
- Sibbald, B and M Roland (1994). “Understanding controlled trials. Why are randomised controlled trials important?” In: *ncbi.nlm.nih.gov* 165, pp. 137–178. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2665449/>.
- Snyder, C. R. (2002). “Hope theory: Rainbows in the mind”. In: *Psychological Inquiry* 13.4, pp. 249–275. ISSN: 1047840X. DOI: 10.1207/S15327965PLI1304\_01.
- Snyder, C. R. et al. (1991). “The will and the ways: Development and validation of an individual-differences measure of hope.” In: *Journal of Personality and Social Psychology* 60.4, pp. 570–585. ISSN: 0022-3514. DOI: 10.1037//0022-3514.60.4.570.
- Snyder, CR et al. (2000). “Hope theory: Updating a common process for psychological change.” In: URL: <https://psycnet.apa.org/record/2001-00353-007>.
- Steiger, James H. (1990). “Structural Model Evaluation and Modification: An Interval Estimation Approach”. In: *Multivariate Behavioral Research* 25.2, pp. 173–180. ISSN: 15327906. DOI: 10.1207/s15327906mbr2502\_4.
- Tanguy, Bernard et al. (2014). “The Future in Mind: Aspirations and Forward-Looking Behaviour in Rural Ethiopia”. In: *SSRN Electronic Journal* September. ISSN: 1556-5068. DOI: 10.2139/ssrn.2514590.
- Tanzania, Girls Not Brides*, (2021). Tech. rep. URL: <https://atlas.girlsnotbrides.org/map/tanzania/>.
- Tanzania HIV Impact Survey 2016-2017: Final Report*, (2018). Tech. rep. Dar es Salaam. URL: [https://phia.icap.columbia.edu/wp-content/uploads/2019/06/FINAL-THIS-2016-2017-Final-Report\\_06.21.19\\_for-web\\_TS.pdf](https://phia.icap.columbia.edu/wp-content/uploads/2019/06/FINAL-THIS-2016-2017-Final-Report_06.21.19_for-web_TS.pdf).
- Tavakol, Mohsen and Reg Dennick (2011). “Making sense of Cronbach’s alpha”. In: *International journal of medical education* 2, pp. 53–55. ISSN: 20426372. DOI: 10.5116/ijme.4dfb.8dfd.

- Tennen Howard, Affleck Glenn and Ruth Tennen (2002). “Clipped feathers: The theory and measurement of hope”. In: *Psychological Inquiry* 13.4, pp. 311–317. ISSN: 1047840X.
- Teresi, Jeanne A (2006). “Overview of quantitative measurement methods: Equivalence, invariance, and differential item functioning in health applications”. In: *Medical care* 44.11, S39–S49.
- The World Bank Group. Tanzania Gender Assessment* (2022). Tech. rep. URL: <https://documents1.worldbank.org/curated/en/099155103312251069/pdf/P1760510ca289d0400a40e03f6408826f07.pdf>.
- Thissen D, Steinberg L. Wainer H. (1993). “Detection of differential item functioning using the parameters of item response models. In Differential Item Functioning”. In: URL: 10.1093/0195305191.003.0028.
- Webb, Darren (2007). “Modes of hoping”. In: *History of the Human Sciences* 20.3, pp. 65–83. ISSN: 09526951. DOI: 10.1177/0952695107079335.
- World Bank*, (2021). URL: <https://www.worldbank.org/en/country/tanzania/overview#1>.
- Zanon, Cristian et al. (2016). “An application of item response theory to psychological test development”. In: *Psicologia: Reflexao e Critica* 29.1. ISSN: 16787153. DOI: 10.1186/s41155-016-0040-x. URL: <http://dx.doi.org/10.1186/s41155-016-0040-x>.