University of Cape Town

School of Management Studies

Interviewer Effects In Quantitative Surveys Using A Door-To-Door Approach

By Naadir Soeker (SKRNAA001)

A dissertation submitted in partial fulfilment of the requirements for the award of the Degree of Master of Social Science in Organisational Psychology

School of Management Studies

University of Cape Town

2019

Primary Supervisor: Ines Meyer

COMPULSORY DECLARATION:

This work has not been previously submitted in whole, or in part, for the award of any degree. It is my own work. Each significant contribution to, and quotation in, this dissertation from the work, or works, of other people has been attributed, cited and referenced.

Signed by candidate

Signature:……. Date: 07/02/2020

Acknowledgements

Firstly, I thank my supervisor, Ines Meyer for her guidance and support throughout the Master's programme, and for providing the opportunity to participate in her fascinating research project. I would like to extend my further gratitude to the National Research Foundation for providing funding for me to participate in this programme. I extend my appreciations and acknowledgements to my mother who battled cancer during the year I completed this study, yet still tolerated me living with her during the course of this dissertation. Finally, I wish to thank the University of Cape Town for providing me with the facilities and opportunity to further my education.

Abstract

Interviewers are a principal source of error in quantitative surveys. While surveys are often self-administered (e.g. in online surveys), it is often required to administer these face-to-face. This is the case, for example, in census surveys in low-income areas where there is little internet penetration, like that of the quality-of-life surveys presently being conducted in multiple countries through a residential door-to-door approach (Carr et al., 2018). In such situations, the social interaction between the interviewer conducting the survey and the interviewee is likely to introduce bias into the survey data collected. Interviewer effects (IE) can influence both item non-response and answer quality, i.e., participants not providing the true answer (Harling, et al., 2019). In an attempt to gain more representative data, this study conducted an exploratory analysis on the possible antecedents and consequences of interviewer effects using the Living Wage survey presently being conducted in South Africa, as the study context. To this end, I examine the systematic biasing effects associated with deploying the same group interviewers (n = 10), of the same ethnicity, age, and of equal gender distribution across five sampling areas in Cape Town in a quasi-experimental design (n = 282). This study highlighted that each interviewer is associated with a unique set of systematic bias that varies dependent on the survey item type. Sensitive items requiring respondents to disclose personal information were the most prone to bias, followed by interviewer-referencing and attitudinal items sequentially. Furthermore, this study found that gender differences in the interview had a marginal influence on the attitudes respondents are willing to share. I hope to contribute to an understanding and critical consideration of the antecedents and consequences of deploying human interviewers for collecting quantitative surveys, especially in a context where ethnic, gender and political differences are loaded in social interactions and are likely to contribute to respondents obscuring their responses.

Table of Contents

List of tables

List of Figures

1. Introduction

Interviewers have been demonstrated to be a principal source of error in quantitative surveys (Adida et al., 2016; Davis et al., 2010; Harling et al., 2019, Kish, 1962; Krumpal 2013; Schnell & Kreuter, 2005). While surveys are often self-administered (e.g. in online surveys), it is often required to administer these face-to-face. This is the case, for example, in census surveys in low-income areas where there is little internet penetration, like that of the quality-of-life surveys presently being conducted through a door-to-door approach in multiple countries (Carr et al., 2018). In this context, the social interaction between the interviewer conducting the survey and the interviewee might introduce bias into the survey data collected. Interviewer effects (IE) can influence both item non-response and answer quality, i.e., participants not providing the true answer (Harling et al., 2019).

In the South African context, three issues complicate the collection of self-report data: Due to the low literacy among many low-income workers, participants in the target group are often unable to complete the questionnaire without assistance. Equally so, the multitude of languages spoken in the country make it necessary to use trained interviewers who can serve as translators. Lastly, prior research has shown that participants with low educational levels also have difficulty understanding Likert-type items, which are the primary item type in quality-of-life surveys (Ellorenco, Teng-Calleja, Bertulfo, Clemente, & Menguito, 2019). This makes it impossible to use self-administered questionnaires in parts of the sample where the presence of an interviewer might introduce data distortion, which, in turn, requires a critical engagement in order to understand how to mitigate these. This is especially relevant in certain survey situations where there are large power and status differentials between interviewers – who in this case represent a university – and low-income participants selected for quality-of-life surveys. By examining some of the possible consequences, such as participants providing responses they perceive to be likely to lead to desirable social or material outcomes (Krumpal, 2013), this study seeks to investigate the possible antecedents to interviewer effects (IE) on self-report data collected in low-income door-to-door survey contexts. To this end, a field experiment was conducted where multiple interviewers collected survey data (N = 282) via face-to-face interviews with residents at their homes. This field experiment was conducted under the ambit of the Living Wage study conducted in multiple communities in Cape Town, South Africa, where participants of a multiplicity of cultural backgrounds and living standards were interviewed. The survey data was then analysed in order to investigate whether any item or subscale of the Living Wage measure differed in any systematic way between interviewers.

This study takes a preliminary step towards ascertaining the degree to which IE might arise given conditions where it is highly likely that participants would adjust their responses owing to the presence of the interviewer. Readers may wish to note that this study was carried out in a context where factors such as ethnicity and socioeconomic differences are politically-charged and are likely to influence the way respondents behave in the presence of an interviewer who, in this case, represents a university outside of the local community, and is of an ethnic background that may or may not be similar to that of the respondent. Therefore, interviewers are situated at different 'distances' to participants in terms of gender, ethnicity, locality, and SES (Harling et al., 2019). Both the effects of gender and ethnicity will be discussed in this paper. However, the effects of interviewer locality and SES was not a concern in the present study where the biasing effects of a group of interviewers hired by the university for the Living Wage project is investigated.

Research Question: How do interviewers influence the responses provided by participants in quantitative survey interviews?

2. Literature Review

A degree of ambiguity is likely to be present when confronted with a stranger for the first time who is asking you questions about your living conditions, as is found in the presently discussed study. In these situations, participants are likely to rely on the interviewer for cues on the most appropriate way to respond (Funder, 1995). These cues may be visible on the interviewer (e.g. branding on clothing) or vocalized (idiosyncrasies when speaking; Ellemer, Gilders, & Haslam, 2004). Respondents relying on such cues are likely to adjust their responses, and in some cases, conceal them, depending on the characteristics of their interviewer (Krumpal, 2013). Schnell and Kreuter (2005) posit that bias related to the interviewer stems from two primary sources: personal characteristics in the interviewer-interviewee dyad, and item characteristics. The effects of IE bias can be observed by allowing interviewers to be active across multiple participants in the same sampling areas, and then observing whether participants respond to a given interviewer in some systematic way when compared to other interviewers under the same conditions. When controlling for situational variables between interviewers, any interviewer variance can be attributed to the interviewers themselves. This would result in a systematic IE bias in a given item's (or subscale's, or scale's) responses. Although a great deal of literature analysing item traits is dated (the oldest included dating back over 70 years), face-to-face survey methodology in a wide variety of social and political research still uses the same item types as when it was first introduced (Kish, 1962). The most common example of these are Likert-type attitudinal items, where participants self-report their disposition on a given attitude from a range of numeric values assigned to represent "low" and "high" on that value.

2.1. Theoretical Explanations for Bias

Interviewers may systematically bias response patterns due to variation in their personal characteristics. Past research has highlighted that a great variety of interviewer characteristics are associated with systematically different response patterns. These characteristics include, but are not limited to, age, ethnicity, gender, socioeconomic status (SES), education and personality (West & Blom, 2017). Differentials on these variables in the interviewer-interviewee dyad have also been demonstrated to affect survey responses in a variety of survey contexts, for example, in surveys polling participants on political attitudes (Adida, Ferree, Posner, & Robinson, 2016; Anderson & Silver, 1986), and surveys evaluating disposition to risk-taking behaviour (Rasinski Willis, Baldwin, Yeh, & Lee., 1999).

Various theoretical explanations have been proposed as to why the interviewer alone, or characteristics interacting in the interviewer-interviewee dyad may bias responses. These biases are proposed to be influenced by, among others, social distance or social desirability

theory (Harling et al., 2019). Social distance theories posit that when interviewers and interviewees perceive each other to be similar, then both item response rates and response quality are higher (Tu & Liao, 2007). Social desirability theories posit that respondents are likely to adjust their answers based on what they think the interviewer wants to hear (Krumpal, 2013). There is support for social distance and social desirability theories working in tandem, because a greater degree of social desirability is present where there are differences between interviewer-interviewee ethnicity (Adida et al., 2016). Any biases in such situations tend to be skewed in the direction of respondents seeking to present their own social identity more favourably, in order to maintain an appropriate interaction with their interviewer. Both theories will be considered in this review.

2.1.1. Social distance. Socially perceived differences present in the interview have been empirically shown to influence responses (Harling et al., 2019). 'Social distance' here is operationalized as the number of observable differences between interviewer and interviewee, pertaining to ethnicity, gender, or any observable social cue indicating social class differences (e.g. clothing; Harling et al., 2019). The greater the number of perceived differences between interviewer and interviewee, the greater the social distance present between interviewer and interviewee. Social differences present in the interview have been associated with responses being adjusted by respondents (ibid). With regards to gender, it appears that interviewees trust their interviewer more when they are a member of the same gender: Interviews carried out in same-gender dyads are shown to elicit less response concealment, especially when sensitive items are presented (Davis, Couper, Janz, Caldwell, Resnicow, 2010). Moreover, female respondents were found to be less likely to disclose sensitive attitudes in the presence of male interviewers (Rasinski et al., 1999). In addition to gender differences influencing responses, female interviewers across interview settings are likely to be perceived as more approachable when compared to males (West & Blom, 2017), and therefore influence the attitudes participants are willing to share. Rasinksi et al. (1999) conclude that the reason gender differences contribute to response concealment is that respondents perceive risks associated with disclosing sensitive information to individuals of greater social distance. In this view, answering a question honestly may be seen as participants making a rational valuation of risks and losses associated with truthful reporting (Krumpal, 2013).

Social distance may also be applied to ethnic differences in the interview. Ethnic differences present in the interview have substantial support for influencing survey responses. For example, respondents interviewed by members of another ethnic group almost always provided socially desirable responses in order to present themselves or their ethnic group in a more favourable light to the interviewer (Stocké, 2007). Respondents have been shown to alter political opinions when the interviewer is of another ethnicity in South African (Adida et al.,

2016) and North American contexts (Anderson & Silver, 1986). In North American contexts, respondents were more likely to disclose political non-participation to interviewers of the same ethnicity (Anderson & Silver, 1986) when compared to attitudes expressed with interviewers of different ethnicities. In Uganda, when the interviewer is of a different ethnicity, respondents are less likely to report that they are politically disadvantaged in order to present their own group in a favourable light (Carlson, 2015). Taken together, these findings suggest that respondents are reluctant to portray their own ethnic group in a negative light when interviewed by an interviewer of a different ethnicity, and would therefore conceal attitudes they deem unfavourable to share. It follows from this analysis of social distance literature that,

H_1: Gender dyad configurations in the interview achieve systematically different results from one another on survey items.

H_2: Respondents of different ethnicities to interviewers achieve systematically different results on survey items to respondents of the same ethnicity to interviewers.

2.1.2. Social desirability theories. Stocké (2007) posits that social desirability in the interview is a precondition that necessitates IE bias. Social Identity Theory (SIT) posits a framework in which social desirability might function in the social world. In SIT, people in social settings are constantly engaged with categorising themselves and others according to a level of social desirability; they subsequently adjust their behaviour in order to align their outward identity with that of a desired group. The opposite is also true, where people in social situations emphasise their distinctness from an undesirable group. SIT can be applied to survey data collection in the following way: if an interviewee perceives an interviewer's outward presentation to represent membership of a desirable group (e.g. their clothing is associated with a desirable university or institution), they are more likely to emphasise opinions that they perceive to be in congruence with those held by members of the desirable group. In other words, the interviewer in this case would elicit a certain social identity expression from the interviewee, thus biasing their responses (Ellemers, Gilder, & Haslam, 2004). Given that interviewers in surveys tend to maintain the same appearance and social traits over the course of multiple interviews, they are likely to produce a consistent biasing effect on the responses they elicit from the participant (Ellemers 1993). This is because they possess relatively invariable social cues to be re-used by several participants, which leads to a consistency in the way several participants respond to them.

There is further evidence to suggest that respondents engage in such impression management consciously, oftentimes in order to gain a perceived material benefit associated with social group congruence (Krumpal, 2013). Rational Choice Theory (RCT) posits that

people, when faced with a multiple-choice situation, respond in ways in order to maximise potential gains and to minimise potential losses (Rasinski et al, 1999). For example, interviewers might be perceived as sources of potential gain to interviewees as 'affluent researchers' with resources to aid (which might occur when university interviewers are deployed in areas of economic need). Respondents obtained from people who engage in this impression management strategy is likely to be biased. This impression management strategy employed by respondents could skew results on a survey item in a 'positive' or 'negative' direction, depending on how the item content is perceived to be viewed by the desirable group (Krumpal, 2013).

For reasons presently unknown, individual interviewers of the same ethnicity and gender groups remain likely to differ in responses, likely because individual respondents vary in their perception of interviewer traits (Ellemers, Gilder, & Haslam, 2004). Respondents may be argued to be prone to social desirability consciously or unconsciously to varying degrees. However, in situations where ethnicity, gender, and attire are kept consistent, the subtle cues that contribute to such variance in respondent perceptions between interviewers are likely remain consistent (Schnell & Kreuter, 2005). These cues might be associated with, for example, interviewer accent, perceived political similarity, perceived education, socioeconomic similarity (Adida et al., 2016), and so on. Consequentially,

H_3: Interviewers systematically differ in the responses obtained from all items.

2.2. Item Characteristics

The degree to which interviewers effect responses vary dependent on item type. In order to calculate IE bias in the present study, each item has to be considered individually, given that different item types vary in their susceptibility to IE bias. Previous studies reveal that IE bias increases for items that are; (1) attitudinal (Krumpal, 2013), (2) sensitive in content (ibid), and (3) contain wording which refer to the interview situation where the question is being asked (Schnell & Kreuter, 2005).

2.2.1. Attitudinal items. Attitudinal items have long been relied upon as a practical tool to gather large-scale observations on self-reported opinions in the social world, yet they present a large number of challenges - chief among them being whether respondents have provided truthful and accurate information (Adida et al., 2016). An attitudinal item mighttake on, for example, a Likert-type format, a yes/no format, or a sliding scale, all of which seek to quantify certain self-report opinions from the interviewee's experience (Kish, 1962). Responses to attitudinal questions are likely to be influenced by perceptions of social desirability across a wide variety of topics (Hogg & Turner, 1985). These include, for example,

political attitudes (Adida et al., 2016), opinions on cultural and ethnic groups, and opinions on psychological states (Krumpal, 2013). In cases where participants have to share attitudes on unfamiliar topics, they are more likely on cues and explanations from the interviewer in order establish the appropriateness of their response (Krumpal, 2013).

Furthermore, attitudinal items tend to take comparatively longer to complete than factual items (for example, items requiring the participant to provide their age), giving participants more time to detect social cues from the interviewer (such as body language approval) which might sway their response (Funder, 1995). Such items inherently reference an attitude that participants might feel socially pressured to obscure or over-represent given the social pressures associated with providing a truthful answer. IE bias can be observed in these items if large groups of people respond to a certain interviewer in a significantly positive, negative, or neutral direction, when compared to other interviewers.

H_4: Interviewers systematically differ in responses obtained on attitudinal items.

2.2.2. Items referencing the interviewer. When survey items contain wording that reference the interviewer in any way, they have been shown to be associated with a degree of bias (Krumpal., 2013). These might include, for example, an item related to appraising an interviewer's performance. In such cases, respondents might try to maintain an appropriate interaction with interviewers by avoiding responses that may be perceived to be abrupt or offensive (O'Muircheartaigh & Campaneli, 1998). In other words, respondents tend to engage in impression management in order to maintain an appropriate interaction with the interviewer, sometimes altering their responses to do so (see section 2.1). Similarly, Hermann (1983; as cited in Schnell & Kreuter, 2005) found smaller IE bias if items were not pertaining to the interviewer. Conversely, larger IE is observed for items where the content is not important to the respondent (ibid). Items that refer to the interview situation highlights the social distance between interviewer and respondent, encouraging the respondent to answer in ways viewed as socially appropriate (Harling et al., 2019).

In the present study, items that reference the interviewer are also attitudinal (see table 2). These items are therefore suspected to be prone to bias for two empirically-supported reasons, in that they reference the interviewer in content and they are attitudinal in format. Therefore, it is predicted that IE is more likely to be prevalent in items that reference the interviewer than in attitudinal items that do not. The consequential hypotheses are as follows:

H_5: Interviewers systematically differ in responses obtained on items that reference the interviewer.

H_6: IE is more prevalent for items that contain wording that refers to the interviewer than attitudinal items.

2.2.3. Sensitive items. Items that probe for sensitive information from interviewees are more prone to biasing effects by the presence of the interviewer (Harling et al., 2019; Krumpal, 2013). Tourangeau and Yan (2007) outline three dimensions of a sensitive item: intrusiveness, threat of disclosure, and social desirability. Firstly, intrusiveness pertains to questions that elicit potentially taboo or intrusive information related to the interviewee's privacy. In these cases, participants are more or less likely to disclose their attitudes on these items depending on the interviewer asking the question. Secondly, threat of disclosure pertains to respondents' concerns about perceived possible risks of negative consequences associated with truthfully reporting a socially perceived taboo position (e.g. job loss or illegal behaviour). Threat of disclosure increases the unlikelihood that a participant discloses sensitive information should there be a perceived likelihood that any responses might become known to a third-party in the larger survey setting. Thirdly, social desirability applies to a sensitive question where answering truthfully is associated with answering socially undesirability. In other words, participants tend to mask sensitive attitudes in the presence of an interviewer in order to manage perceived social pressure (Stocké, 2007). In fact, even in the absence of an interviewer (for example, participants given the option to respond to a ballot box anonymously), respondents are reluctant to disclose sensitive political attitudes (Langhaug, 2011).

Sensitive items are associated with a higher nonresponse rate (Stocké, 2007). In household income measures, similar to the one deployed in the present study, nonresponse rates of sensitive items are between 20% and 26.2% (Krumpal, 2013). Harling et al. (2019) find empirical support that interviewer-respondent dyads with greater social distance are more prone to bias in sensitive than other item types. Taken together, these findings support that items sensitive in content are associated with a greater degree of bias than other items.

As a whole, differences in interviewer characteristics are likely to compound present IE found in various item types. For example, when sensitive items are being presented, IE is more likely to become distinguishable where there are ethnic differences present in the interview (Davis et al., 2010; Harling et al., 2019). If this this effect is observed in the present study, the highest degree of IE will be observed when sensitive items are being presented while there are ethnic and gender differences present in the interview (Krumpal, 2013). This is because sensitive items are the most prone to IE (Harling et al., 2019). By contrast, the same social differences present in an interview are likely to be masked on scores obtained on attitudinal items (Krumpal, 2013) and items referencing the interviewer (O'Muircheartaigh &

Campaneli, 1998), which are comparatively less prone to IE. Consequentially:

$H_{7:}$ Interviewers systematically differ in responses obtained in sensitive items.
H_8: IE bias is more prevalent in sensitive items than in other items.

However, it may also be that each interviewer differs across all items regardless of social differences (thereby providing contrasting evidence to predictions related to ethnicity and gender). This is supported by literature finding support that interviewers elicit social desirability differing degrees owing to idiosyncrasies in their presentation of the same items (Ellemers, 1993, Ellemers, Gilder, & Haslam, 2004). In the present study, IE will be analysed by gender and ethnic groupings present in the interview, and then by interviewers individually across each item type.

3. Method

3.1. Research Design

This study employed a quasi-experimental design, where multiple interviewers were deployed to collect data in the same five residential areas. Quantitative data was gathered using a self-report, interviewer-administered survey. This questionnaire was distributed in hard-copy to randomly-selected participants at their homes. An interviewer-feedback section was provided at the end of the survey where participants described their experience with the interviewer. Interviewers ethnicity was controlled; an even number of males and female interviewers were deployed to conduct interviews in the same areas. The effect each might have on the data collected was analysed through statistical procedures after data collection.

3.2. Participants and Sampling

In the present research, a simple random sample may not be used to generalise findings to the entire working class South African population, owing to practical constraints related to transport costs and nationwide interviewer availability. Instead, the present research used a clustered sample (n = 282) from five randomly-selected wards in Cape Town. Cape Town's wards were selected using WaziMap (https://wazimap.co.za/), an online tool designed by Cape Town-based NGO OpenUp showing the demarcations of Cape Town's municipalities, or 'wards'. These areas were selected at random using the Cape Town municipal area as a cut-off. What follows is a description of the conditions of each ward.

Areas where the military was intervening due to an exceedingly high violent-crime rate was excluded from our ward selection due to safety concerns. 40 cases were removed from the final analyses owing to them being interviewed by interviewers who did not complete sufficient cases for subsequent analyses (n < 13); This removal was based on whether there were sufficient cases per interviewer for chi-square analysis of cluster membership by interviewer (section 4.2.2). A total sample size of n = 282 cases were retained for analysis. Once wards were selected, interviewers approached houses in these areas in one walking shift during weekend mornings (9.30 a.m.) until afternoons (14:30 p.m.) owing to feasibility and participant availability constraints.

Wards vary in shape and size owing to a variety of factors, including interdependence of people, existing patterns of human settlement, employment, transport movements, spending, access to municipal services and other qualitatively meaningful characteristics (Municipal Demarcation Board, 2016). Given that wards in Cape Town were drawn through a qualitative committee decision, they are not standardized related to housing format and density. The result is that each ward selected had varying densities and distributions of housing. Each ward was greatly variant in size - some being over ten square kilometres (e.g. Cape Town Ward 58) and some being under two square kilometres (e.g. Cape Town Ward 47). As such, we

subdivided wards into residential areas by looking at the map in order to see where there were at least six blocks of housing. Wards that consisted primarily of unhoused area were replaced (e.g., Cape Town Ward 32). Wards that met these criteria were selected for our interviewers to walk through systematically, starting from the north-western most block, of the northern- most residential area. Interviewers made their way South and East, until the walking shift had elapsed.

Every fifth house was selected for potential participation as interviewers walked. If the initially-selected house is a nonresponse, this house would be replaced with the one nearest or adjacent to it, up to five times. If the replacement house was a successful response, interviewers moved on to the next fifth house relative to the house that participated. If all five houses were nonresponsive, then the next set of five houses was selected within a ward. This continued until the data collection shift had elapsed, with the goal being to complete as many surveys as possible within a shift. A standardized quota for participation per ward could not be set owing to time and feasibility constraints when employing data collectors on weekends.

The first two wards included (Cape Town municipal wards 102 and 111) were working-middle-class suburban areas set next to a highway in the Northern Suburbs of Cape Town. Almost all houses in these wards were single-story structures, with some kind of fencing marking the perimeter. Most of the fencings of these houses were gate-accessible, and did not have a bell outside the gate, so interviewers were required to enter the premises of the front porch in order to solicit attention from the resident by ringing a doorbell or knocking on the front door. In the mornings, these areas did not have any residents outside barring few exceptions. As the day reached lunchtime (1 p.m.), some streets of these wards were marked by residents and children either gathering or playing outside in the afternoons. The streets of these suburban wards were unidirectional, with little shade or trees and houses lined up shoulder-to-shoulder on each side of the street.

The remaining wards (34, 41, and 94) were in township areas of Cape Town. These wards had main streets with more densely-arranged semi-formal urban housing. These houses were sometimes made of bricks (sometimes including a fenced porch), and sometimes they were an informal shack homestead made of steel sheets and wood. In general, these areas had a main street with avenues that lead off the main street to narrower lanes. Both the main streets and the lanes were dotted with tightly-shouldered residential housing, though the main streets tended to be busier with foot and vehicle traffic. Township areas had more residents out in the streets at all times of day when compared to other wards. Groups of residents would gather outside, and often times express active curiosity in the uniformed interviewers walking the streets; they were more likely inquire about the study as interviewers passed, allowing for an easier request for participation if they were from a household included for participation. Streets

in these wards would vary by how busy they were, and also in their levels of safety. Interviewers would solicit advice from residents about which streets to avoid owing to their having been a history of unrest in these areas, and then plan to walk another street for safety concerns.

3.3. Procedure

For interviewer training, nineteen research assistants were trained to assist with data collection. Of these, six were removed due to having too few completed cases for analysis, as required by Chi-squared analysis ($n < 13$). Three more interviewers were removed due to them being irrelevant to this study, for example, students hired through the university who collected data in other areas. These were excluded as their data was to be collected from other sampling areas, and consequentially and could not be analysed for interviewer effects without controlling for systematic variance introduced by sampling area. Following the removal of interviewers not relevant to the present study, ten interviewers remained. They were each arbitrarily assigned an 'interviewer number' between 2 and 13. The remaining interviewers were unemployed Xhosa and English bilingual speakers from community youth group in Cape Town. These were five male and five female interviewers, all of ages between 18 and 20 years old.

Interviewer training served to ensure that research assistants were familiar with the study, its aims and procedure as well as with the survey instrument. During the interviewer training, the entire survey was walked through with interviewers question-by-question to check for a mutual understanding of all survey items. The training also workshopped approaching strangers at their houses through role-playing. Interviewers also practiced a greeting sequence, which included (1) introducing the aims of the study, (2) informing participants about the study incentive and then (3) requesting participation. While the idiosyncrasies of how interviewers approached doing this sequence introduced slight differences, this greeting sequence was to be kept consistent. Interviewers were made aware that my study was interested in investigating bias and therefore included this interviewer training session and the interviewer feedback section (discussed in section 3.4). Interviewers were also told that the university was interested in reducing survey bias and therefore included the interviewer feedback section. Finally, the training covered logistical concerns associated walking through residential areas on the weekends, such as safety, weather conditions, fatigue concerns, transport and meeting times.

After attaining faculty ethics approval, the group of interviewers entered into a residential ward on a Saturday, as it was the day on which working individuals were most likely to be at their homes. After a resident was approached by knocking on the door or ringing the bell, interviewers requested participation with the selected house by garnering interest for the study and its aims. Residents were then offered an incentive to participate in the study, which was

a ticket to participate in a lucky draw to win one of three R500 cash prizes. All research assistants wore the same uniform T- shirt, which was yellow in colour, identifying them as members of the University of Cape Town. Interviews would take between twenty and thirty minutes to complete. Research assistants recorded their own name and ward number on the front of the sheet. At the end of the interview, participants were given the opportunity to record any additional comments they might have in the 'feedback' section of the survey.

Participation in the study was voluntary and anonymous. Once a participant has agreed to consider participating in the research, the research assistant provided a brief overview of the study and proceeded to ask the survey questions to the participant. In this way, individuals with low English literacy were able to partake in the questionnaire. The questionnaire was available in English and Afrikaans, with the research assistants being able to serve as translators for other languages. If the participant was not available at the time of being approached, then research assistants asked if they would be willing to complete the questionnaire by themselves. When an survey was dropped off in this way, a note was made that the survey was conducted without an interviewer. In this way, participants would not be excluded if they showed interest in the study but were presently preoccupied. These cases were removed from IE analyses, owing to the fact that these participants are exposed to interviewers to a comparatively shorter length of time, potentially confounding the research design. Interviewers agreed to do the survey where the respondent felt most comfortable at their home. All completed surveys were captured and stored electronically using IBM SPSS statistics version 23.

3.4. Measures

A shortened version of the Living Wage and Capabilities measure (LWC; Ellorenco et al., 2019) was administered, the final version containing 87 items (Appendix B). The 22-item Organisational Citizenship Behaviour and Task Performance (OCB-TP; Williams & Anderson, 1991) subscale of the LWC was analysed in is study, situated at block 4 of the larger LWC survey (after item 42). This subscale begins after participants were asked about perceived freedom to attain specific domains of life (block 3). This subscale was of interest for this study because it elicited attitudes in a self-report Likert-type format from participants that could be used to analyse for IE (discussed in section 4). Likert-type response scales which range from "strongly disagree" to "strongly agree" provide a conceptual challenge for low-income workers (Carr et al., 2018). Instead, most items are phrased in such a way that they can be answered using frequency indications (e.g. never, almost never, sometimes, almost always, always).

Three additional items were added at the end of the OCB-TP. These items requested that participants rate their experience with the interviewer, and the interviewer's physical appearance (items 85 to 87, appendix B). This section was presented as feedback the university

might be interested in, in the form of a rating of the interviewer's performance. Participants were given two Likert-type items (items 85 and 86) to rate interview experience and physical appearance respectively, and a section to share their own opinion (item 87) about the interviewer with any additional comments.

The full Living Wage Survey measure (Appendix B) consisted of 87 items and took an average of 20 minutes to complete. The responses of ten items were included for IE bias analyses as they met both the theoretical criteria for response bias (i.e. attitudinal or sensitive items; Krumpal, 2013) and were of the appropriate response format – Likert type scales – which allowed for analysis of any systematic answer trends. The resultant IE analyses included five attitudinal items (coded as items one to five), two items that reference the interviewer (coded as items six and seven), and two items marked as being 'sensitive', as they contain wording that pertain to attitudes related to the respondent's income (coded items eight and nine). One attitudinal item was later removed due to having too few completed responses. See table 1 for a description of items included in this study.

Table 1

Summary of Items Included for IE Analysis

Item Type	Item code	Measurement Scale	Response Categories	Item description
Attitudinal items	1	Continuous	*Range from 1 (Never) to 6 (Always)*	Mean derived from the six items of the 'altruism' subscale - OCB-TP (items 42 to 48)
	2	Continuous	*Range from 1 (Never) to 6 (Always)*	Mean derived from the four items of the 'conscientiousness' subscale of the OCB-TP (items 51 to 56)
	3	Continuous	*Range from 1 (Never) to 6 (Always)*	Mean derived from the six items of the 'job satisfaction' subscale of the OCB-TP (items 60 to 63)
	4	Ordinal – measured on a six-point Likert scale.	*Never, almost never, Sometimes, often, almost always, always.*	Item 65: "I am satisfied with my job in general"
	5	Ordinal - measured on a six-point Likert scale.	*Never, almost never, Sometimes, often, almost always, always.*	Item 70: "I am satisfied with life in general"
Items Referencing Interviewer	6	Ordinal – measured on a seven-point Likert scale.	*Very poor, poor, quite poor, neutral, quite good, good, very good*	Item 85: "Overall, how would you rate your experience with your interviewer?"
	7	Ordinal – measured on a seven-point Likert scale	*Very poor, poor, quite poor, neutral, quite good, good, very good*	Item 86: "Overall, how would you rate the physical appearance of your interviewer?"
Sensitive Items	8	Ordinal - measured on a six-point Likert scale.	*Never, almost never, Sometimes, often, almost always, always.*	Item 66: " I am satisfied with my pay"

9	Ordinal - measured on a seven-point Likert scale.	*Very unfair, unfair, quite unfair, neither fair nor unfair, quite fair, fair, very fair.*	Item 71: "Overall, how fair would you say your wage is?"

3.5. Statistical analyses

IE was assessed by item. This was because each item had a unique response distribution by interviewer. Score variances were then analysed in accordance to item type (i.e. attitudinal, interviewer-referencing or sensitive) in light of literature proposing varying degrees of bias for each item type. When an item type had three items or more (as is the case in attitudinal items in this study), two-step cluster analysis was used in order to assess whether responses were distributed in any systematic way, as cluster analysis requires a minimum of three items to be run (Rapkin, & Luke 1993). When individual interviewer differences were a variable of interest, responses within cluster memberships were cross- tabulated by interviewer, and chi-squared tests were conducted in order to examine whether responses significantly differed from expected response per cluster membership. To further investigate which interviewers influenced cluster membership, a z-distribution significance test was done to ascertain whether an interviewer's responses differed significantly from the expected distribution of responses according to the cluster model obtained.

When item types had two or less items, analysis-of-variance (ANOVA; in the case of scale measures) or Kruskal-Wallis tests (in the case of ordinal measures) were done. Groupings were made by interviewers individually by or gender/ethnicity groupings, in order to test whether interviewers, or any grouping of interest, differed from one another on a given item. Post-hoc analyses test were used to determine where individual differences in responses lie in the results of each test.

4. Results

This section of the paper is concerned with testing the measurement properties of the study subscale, and the study hypotheses stated in the literature review First, measurement properties of an included study scale are presented. Following this, hypotheses related to testing differences between participant-interviewer dyads are tested. Then, individual differences per interviewer on every studied item (n = 9) are analysed. Finally, scores obtained on physical appearance appraisals per interviewer are investigated.

4.1. Reliability and Validity Analyses

The OCB-TP subscale of the Living Wage Survey consisted of 22 items and was found to be highly reliable (a = .78), where Cronbach's alpha values above .70 are found to be sufficiently reliable (Steiner, 2003).

An exploratory factor analysis (EFA) using principal axis factoring with direct oblimin rotation was conducted on the OCB-TP scale in order to determine the validity. This was done so that subsequent IE analyses could be summarized into one or a few average scores to be used in further analyses. Furthermore, principal axis factoring allowed to examine whether this scale extracted any potential factors in our sample, in order to examine whether scale means represent any potential meaningful dimension extrapolated from the data in our sample. This would allow for examination into potential IE biases at the construct level.

Items were excluded from the final version of the scale if they met one or both of two criteria: whether they had factor loadings below .30 or whether they cross-loaded across multiple factors too significantly to be considered independent from one another. Items with factor loadings more than .30 represented items that accounted for a sufficient amount of variance in the underlying construct (Young & Pearce, 2013). When differences between loadings on the factors fell below .25, they were considered too close to represent meaningfully different variables. The cross-loading of one item with a loading above .30 across multiple factors was undesirable because it represents an item measuring multiple dimensions, despite it being designed to measure one factor exclusively with a high factor loading in its intended population.

A first iteration of the factor analysis revealed four factors, with two items cross-loading across multiple factors (r > .30), and an additional two items loading poorly onto a fourth factor (see Table 1, Appendix A). These four items were removed from a second iteration (see Table 2, Appendix A), revealing that two more items cross-loaded too strongly (r > .30) and were

subsequently removed. The final iteration (Table 2) contained 16 items, each with factor loadings greater than .30 loading onto one of three factors. Item memberships to their respective subscales were consistent with the initially proposed OCB-TP scale (Williams & Anderson, 1991) after the abovementioned six complex items were removed from the OCB-TP subscale. Factor one consisted of six items, with factor loadings between .70 and .81. Factor two consisted of six items, with factor loadings between .50 and .81 Factor three consisted of four items, with factor loadings between .55 and .91; factor loadings greater than .4 are considered sufficient for inclusion by conventional standards.

Table 2
Results from Factor Analysis of the Organizational Citizenship Behaviour and Task Performance (OCB-TP) Scale

OCB-TP item	Eigenvalue	% explained variance	Factor loading
Factor 1: Altruism	4.17	26.03	
42. I help others who have been absent from work.			.67
43. I help others who have heavy workloads.			.81
44. I volunteer to assist supervisors with their work.			.73
46. I go out of my way to help new employees.			.63
47. I show concern to my co-workers.			.62
48. I share information with co-workers.			.50
Factor 2: Conscientiousness	3.93	24.57	
51. I take work breaks that are not permitted. (R)			.73
52. I spend much time with personal phone. conversations. (R)			.81
53. I complain about unimportant things at work. (R)			.70
55. I only put minimal effort into my current job because I really feel that I'm wasting my time at work. (R)			.84
56. I don't know why I'm doing this work; it's pointless work. (R)			.69
56. I do little work in my current job because I don't think this work is worth putting a great deal of effort into. (R)			.73
Factor 3: Job Satisfaction	1.83	11.46	
57. I put a great deal of effort into my current job because the work I do is interesting.			.55
61. Overall, I am content with my work.			.80
62. I am content with the important aspects of my job.			.91
63. I like the work that I do in my job.			.70

Note. N = 282. The extraction method was principal axis factoring. Rotation Method: Oblimin with Kaiser Normalization. Factor loadings below .30 removed. Reverse-scored items denoted with an (R).

The three factors were labelled 'altruism' (6 items, a = .84), 'conscientiousness' (6 items, a = .90) and 'job satisfaction' (4 items, a = .81) respectively. The wordings of items were investigated to confirm what construct each of the three subscales represented; in each subscale, the wording of items aligned with the factors found in Williams and Anderson's (1991) original research. For each participant, a mean per subscale was formed to indicate their self-described

levels of altruism, conscientiousness and job satisfaction. These three scores were per participant were included as 'attitudinal' items in further analyses.

4.2 Results Pertaining to Study Hypotheses

4.2.1 Gender and ethnicity. Hypotheses H_1 and H_2 pertained to interviewer-interviewee gender and ethnic differences respectively. H_1 stated that gender dyad configurations in the interview achieve systematically different results from one another on survey items. Four gendered dyads were possible in our interview situation given male or female interviewers and male or female respondents; these four dyad configurations were compared through ANOVA. Dyads are coded for ease of reporting in Table 3. H_2 Stated that respondents of different ethnicities to interviewers would achieve systematically different results on survey items compared to respondents who shared the same ethnicity with interviewers. To this end, scores of respondents of varying ethnicities were compared to see whether there were significant differences between them. Given that all interviewers were Black African, interviewers of other ethnicities are assumed to be of a 'further' social distance, whereas respondents who were Black African are assumed to be of relatively 'nearer' social distance. The scores between Black African respondents and respondents of other ethnicities were compared through ANOVA.

Table 3
Interview-Respondent Gender Dyads Codes

Interview Condition	Dyad Code	*n*
Males interviewing males	1	68
Males interviewing females	2	67
Females interviewing males	3	70
Females interviewing females	4	77

Same-gender dyads were considered to be of 'nearer' social distance whereas different-gender dyads were considered to be of 'further' social distance. Furthermore, dyads of males interviewing female respondents are assumed to be of the 'greatest' social distance given that females are empirically found to be more approachable in a variety of interview contexts (Dykema, Diloreto, Price, White, & Schaeffer, 2012; Pollner, 1998). Therefore, the 'females interviewing males' condition (dyad 3) is assumed to be less prone to bias than the 'males interviewing females' condition (dyad 2). Cases where respondents indicated 'other' or 'prefer not to say' were omitted from this analysis owing to the fact that they constituted a marginal amount of data (n = 4). Ethnicity was controlled when conducting ANOVA on gender groupings in order to test the effect of gender in isolation; consequentially, only the 'Black African' ethnicity (n = 222) was included for gender bias analysis.

Gender. H_1 asserted that gender differences in the interview situation influenced the scores obtained from the interview. An ANOVA was conducted comparing the four interviewer-respondent gender dyads across the three OCB-TP subscales, only including participants who indicated 'Black African' as their ethnicity (n = 222). ANOVA was selected to test this hypothesis because any significant differences in scores between gender dyads on an item would indicate that those scores systematically differed depending on the gender of interviewer and interviewee.

Levene's test for homogeneity of variance was insignificant for all three subscales, namely Altruism [$F_{(3, 218)} = .89$, $p = .45$], Conscientiousness [$F_{(3, 217)} = 1.03$, $p = .382$], and Job Satisfaction [$F_{(3, 214)} = 1.46$, $p = .23$]; therefore ANOVA may be used to analyse for any significant differences in variance between the gender dyads. ANOVA was significant for Altruism [$F_{(3, 218)} = 3.02$, $p < .05$] and Conscientiousness [$F_{(3, 217)} = 3.95$, $p < .05$], and nonsignificant for Job Satisfaction [$F_{(3, 214)} = 2.00$, $p = .11$]. Post hoc comparisons of gender dyads using the Tukey HSD test (See Table 3, Appendix A) show that dyad three significantly differed from dyads one and two on both Altruism and Conscientiousness ($p < .05$). Furthermore, dyads two and three significantly differed from each other for both Altruism (dyad 2, M = 4.44, SD = 1.06; dyad 3, M = 4.83, SD = 1.02) and Conscientiousness (dyad 2, M = 4.28, SD = .96; dyad 3, M = 3.60, SD = 1.50). Taken together, these findings support that interview conditions where females interview males achieve responses that significantly differ to several other interview conditions. Furthermore, different-gender dyads achieve differing results dependent on whether male or female interviewers are deployed. H_1 asserting that gender differences in the interview influence IE is supported.

The effects of gender in the interview were further assessed on items marked as 'sensitive items' (items 8 and 9). This was because biases associated with gender have been observed sensitive items (Krumpal, 2013). These items required participants to reveal sensitive information related to their attitudes on their pay (Rasinski et al., 1999). Sensitive items in our sample were measured via ordinal scales; therefore, a Kruskal-Wallis test was used to compare response patterns between gender dyads in order to ascertain whether there were any significant differences in response patterns between these groups. A significant Kruskal-Wallis test indicates that there were systematic differences in the way respondents responded to interviewers (H_1). Kruskal-Wallis tests were insignificant for both item eight [$H_{(3)} = .28$, $p = .42$] and nine [$H_{(3)} = .11$, $p = .99$]. These findings suggest that the gender configuration of the interview had no biasing effect on the responses obtained on sensitive items.

Ethnicity. H_2 asserted that ethnic differences between interviewers and interviewees affected the scores obtained on survey items. To ascertain whether this was the case, ANOVA was conducted comparing respondent ethnic groups in scores obtained across the three OCB-TP subscales. ANOVA was selected to test this hypothesis as any significant differences between ethnic groups would indicate that survey responses differed depending on respondent ethnicity. In total, four ethnicity groups were included for analysis, namely 'Black African' (n = 222), 'Coloured' (n = 21), 'White' (n = 9), and 'Other' (n = 7). Three ethnicity categories were removed from analysis for having too few valid cases ($n < 4$). These were, 'Asian' (n = 1), 'Indian' (n = 2), and 'prefer not to say' (n = 3). Readers may note that that the number of participants in the ethnic groups that were included varied greatly, which might have contributed to an exaggeration of any differences in score variance between the ethnic groupings (Lix, Keselman & Keselman, 1996). Results of difference tests between ethnic groupings need to be interpreted with extreme caution should any differences between groups be found.

Levene's test for homogeneity of variance was nonsignificant for all three subscales, namely Altruism [$F_{(3,255)} = .10$ $p = .96$], Conscientiousness [$F_{(3,254)} = 2.48$, $p = .06$], and Job Satisfaction [$F_{(3,\ 251)} = 1.09$, $p = .35$], therefore ANOVA could be used to analyse for any significant differences in variance between ethnicity groupings for these items. Table 4 summarises the mean, minimum, maximum, standard deviation and 95% confidence interval of the mean of ethnic groups on the three OCB-TP items. ANOVA was nonsignificant for all three subscales, namely Altruism [$F_{(3,255)} = 1.63$, $p = .184$], Conscientiousness [$F_{(3,254)} = .281$, $p = .839$], and Job Satisfaction [$F_{(3,251)} = .722$, $p = .54$]. Taken together, these results indicate that ethnic differences or similarities between interviewers and interviewees in our sample did not relate to any systematic differences in scores obtained for these items. Therefore, H_2 asserting that ethnic differences in the interview influence responses has no support.

Table 4
Descriptive Statistics Per Ethnicity Grouping Items One to Three

						95% Confidence Interval for Mean	
Ethnic Group	*n*	*M*	*SD*	Min	Max	Lower Bound	Upper Bound
Item one: 'Altruism'							
Black African	222	4.65	1.01	1.71	6.00	4.51	4.78
Coloured	21	4.83	.96	3.29	6.00	4.39	5.26
White	9	4.74	1.29	2.00	6.00	3.75	5.73
Other	7	3.85	1.48	1.00	6.00	2.47	5.21
Total	259	4.64	1.04	1.00	6.00	4.52	4.76
Item two: 'Conscientiousness'							

Black African	221	3.96	1.08	1.33	5.67	3.81	4.10
Coloured	21	4.16	1.11	2.33	5.83	3.65	4.66
White	9	4.12	1.05	2.00	5.17	3.31	4.93
Other	7	3.93	.39	3.50	4.33	3.65	4.29
Total	258	3.98	1.07	1.33	5.83	3.85	4.11
Item three: 'Job Satisfaction'							
Black African	218	5.00	.94	1.50	6.00	4.86	5.12
Coloured	21	5.00	.95	3.00	6.00	4.58	5.43
White	9	5.33	.70	4.00	6.00	4.80	5.86
Other	7	4.64	.71	3.75	6.00	3.98	5.31
Total	255	5.00	.93	1.50	6.00	4.88	5.11

4.2.2. Interviewer differences across all items. H_3 stated that participant responses significantly differed between individual interviewers when data is collected under the same conditions. In conditions where participants were randomly distributed between interviewers, one way to ascertain bias in responses is to examine whether the response patterns of participants are distributed into distinct groupings, or 'clusters'. This result also allows for further examination if there is an overrepresentation of gender dyads (H_1) or participant ethnicity (H_2) across any clusters that might emerge from this analysis. If the observed response distribution of an interviewer or any group of interest significantly differs from the expected response distribution modelled through cluster analysis, then responses collected by that interviewer (or group) are likely to be systematically biased (H_3). In other terms, if an interviewer (or group of interest) is exerting a biasing effect on responses, their responses would be overrepresented in one cluster.

To analyse for any response patterns, a two-step cluster analysis was conducted on all analysed study items (n = 9) for which the response format ranged from 1-6 or 1-9 Table 1 (section 3.4) summarises the items included in this analysis. A two-step cluster analysis was used given that it combines hierarchical and non-hierarchical approaches to formulating clusters. The hierarchical component of two-step clustering starts by treating every data point, i.e. every participant's responses across the items of interest, as an individual cluster. Next, data points are grouped with its nearest neighbouring data point; these two data points form a cluster. Clusters are separated when the distance between the next included data point is significantly larger than previous distances, based on one or several distance measurement techniques (see Rapkin & Luke, 1993). This process allows clusters to emerge from data instead of being prescribed by the researcher.

The non-hierarchical component divides the data set into a deliberate pre-determined number of groups, and draws clusters based on where data points more closely fit into one group

over the other, based on each data point's distance from a cluster's centre point, or "centroid" (Rapkin & Luke, 1993). If the number of clusters is not specified in advance, solutions with different numbers of clusters are recalculated and so are their respective centroids, and data points are reassigned cluster memberships based on these recalculated centroids. This process is repeated until the best-fitting non-hierarchical cluster solution is selected for the observed data. The two resulting cluster solutions are then matched to the observed data in order to formulate a parsimonious cluster solution that best fits the observed data based on a set of latent model selection criteria. Included in SPSS and used conventionally in model fitting statistics are Schwarz's Bayesian Criterion (BIC) or Akaike's Information Criterion (AIC).

In practice, one must select a criterion (or set of criteria) when measuring the utility of a cluster model depending on the underlying assumptions about the nature of the 'true' responses present in the population. BIC is a more effective model selection criterion when there is in a reality a salient 'true' model underlying the spread of observed data in the variables of interest (for a more detailed discussion of model selection criteria see Vrieze, 2012). This is because BIC will specify a 'true' model over the observed data spread, whether there is one present or not. In cases where there is no underlying 'true' model governing the spread of observed data, the use of BIC is more likely to lead to error, as BIC is guaranteed to assume a 'true model' in the data spread (Vrieze, 2012). This error increases as sample size increases, as it becomes more likely that parameters that are actually zero will be included as significant. BIC is therefore guaranteed to select a 'true' model as sample size becomes infinitely large. To account for this, BIC model parameters compensate by making statistical significance of the model fit more difficult to achieve as the sample size increases. The null hypothesis being tested by the BIC model selection test is that the 'true' response model implicit to the population differs significantly to the cluster model found through the clustering method. If rejected, the alternative hypothesis supported by BIC is that the model selected and the underlying 'true' response model in the population are the same. At the 300-participant level, which is near to the sample size in the present study, BIC reaches statistical significance at $p = .017$ (Vrieze, 2012). In a sample of 10,000 participants, the BIC requires a significance level of .002 (ibid).

AIC, by contrast, will not impose a 'true' model over the data when one cannot be derived, regardless of sample size. The AIC was derived by Akaike (1987) as an estimate of Kullback-Leibler (K-L) divergence, which aims to measure the distance between any number of proposed ('candidate') cluster models and the 'true' model present in the observed data. AIC assumes that the 'true' model for any spread of data can never be known, and therefore the absolute divergence present between the candidate model and the true model, is unknown.

Instead, AIC fits any number of candidate models over the data and ranks them according to their expected K-L divergence. The candidate model with the lowest expected K-L divergence has the highest model fit to the observed data. Therefore, in cases where there is no 'true' model, AIC is more likely to produce a candidate model with the lowest expected K-L distance, regardless of sample size, with a fixed statistical significance cut-off.

In general, BIC is preferred when there is assumed to be an underlying 'true' model governing the data spread, while AIC is used when the 'true' model for data spread is too complex to estimate parametrically (Shao, 1997). In the present study, cluster models using both BIC and AIC were considered owing to the fact that there is some theoretical basis to assume that the data would be spread in some systematic way, and that there would therefore be an underlying 'true' model underlying the distribution of data in our sample. In the case of the relatively small sample size (n = 282), BIC had a relatively high chance of being statistically significant (and therefore more likely to produce a type-I error when assuming the cluster model fits on a sample), warranting a cross-check of any emerging cluster model produced through BIC with AIC.

In order for cluster analysis to account for any systematic response patterns, cluster quality has to be sufficiently adequate in order for the cluster model to adequately explain the observed spread of data. To test this, the silhouette measure of cluster cohesion and separation is observed to measure how closely the data points lie along the projected cluster model. Cluster quality is measured by two criteria: firstly, how closely data points converge on a single cluster (cohesion); and secondly, how distant clusters are on average apart from one another (separation). A cluster model of high quality will contain data points distributed in a way that is both, highly cohesive and separated. In other words, the cluster model would accurately explain the observed clusters of data. By contrast, a cluster model of low quality will have a cluster solution that is neither cohesive nor separate, i.e. data points would overlap across clusters and not be distributed in tightly cohesive groups. In this case, the cluster model would not provide any useful explanation for how observed data might be distributed (Rapkin & Luke, 1993).

When all 9 items were included in the cluster analysis, the silhouette measure of cluster cohesion and separation was zero (see Figure 1, Appendix A) using AIC and BIC, suggesting that the data did not distribute itself in any cohesive way to warrant further use of the two-step cluster solution for these items. This led to the assumption that clusters may emerge if different item types (i.e. attitudinal items, sensitive items and interviewer-referencing items) were analysed separately.

4.2.2. Hypotheses Relating to Item Types

4.2.2.1. Attitudinal items. *Cluster analysis.* H_4 posited that interviewers obtained results that systematically differed on attitudinal items. In order to test this hypothesis, a two-step cluster analysis was conducted on all attitudinal items (one to five) in order to determine whether responses were distributed to any systematic subsets. Subsequently, responses of interviewers could be compared to any potential response distributions through a chi-square test in order to examine whether any interviewer would be associated with a response trend, and thus suspected to bias the data.

A cluster solution emerged using a two-step clustering method, albeit with low-quality (silhouette measure of cohesion and separation = .1; see Figure 2, Appendix A), where a solution of good quality would be between .3 to 1 (Rapkin & Luke, 1993). This solution was obtained using both BIC and AIC. Figure 3 provides a graphical representation of how survey responses distributed themselves along these clusters by providing the average score for participants in cluster one (n = 144) for the five items and those same average scores for participants in cluster two (n = 108). The low silhouette measure of cluster cohesion and separation denotes that observed responses were not clustered cohesively enough around their respective clusters. The value further denotes that the clusters themselves are on average not distinct enough to be considered meaningfully different, as reflected in figure 3, with the exception of item two. In other words, while two main response patterns emerged in our sample, they were too similar to be considered distinct. Further analysis using this cluster solution was deemed inefficacious in uncovering systematic response patterns.

Figure 1. Means of response clusters derived from attitudinal items.

ANOVA. An alternative way to assess for differences in responses on attitudinal items between interviewers is to compare the average score participants received per interviewer to each other. This would test for any significant difference in response variance between interviewers on each item. This was considered a second resort to cluster analysis because it required multiple tests of significance between each interviewer across each item, and multiple tests of statistical significance are more likely to lead to statistically significant results (Simmons, Nelson & Simonsohn, 2011). It thus increases the chances of making a *type-I* error, that is, a difference between interviewers that is inappropriately accepted though in reality there is no difference between them (Felid & Armenakis, 1974).

A significant difference in participants' item scores found between interviewers would denote that the responses of participants interviewed by different interviewers varied in ways that differed from one another significantly. Conversely, without IE bias, the analysis would be insignificant, supporting no difference in response variance between interviewers. A condition for ANOVA is that the dependent variable is continuous. The three OCB-TP subscale means were continuous, the remaining two items were ordinal. For these items, a Kruskal-Wallis nonparametric test was used, which tests for significant differences between groups on ordinal scales.

Levene's test for homogeneity of variance was significant for Conscientiousness [$F_{(9, 261)} = 4.55$, $p < .01$] and Altruism [$F_{(9, 260)} = 3.85$, $p < .01$], and nonsignificant for Job Satisfaction [$F_{(9, 257)} = 1.81$, $p = .07$]. Therefore, a Brown-Forsythe test was used to measure for any significant difference between groups for items one and two, and an ANOVA was used for item three. The Brown-Forsythe test for equality of means was significant for Altruism [$F_{(9, 212,26)} = 4.71$, $p < .01$] and Conscientiousness [$F_{(9, 230,6)} = 9.31$, $p < .01$]; ANOVA was significant for Job Satisfaction [$F_{(9, 257)} = 3.67$, $p < .01$], suggesting that responses significantly differed between interviewers on all three items, thus supporting the hypothesis. Table 5 summarises the mean, minimum, maximum, standard deviation and 95% confidence interval of the mean of each interviewer on the three OCB-TP items.

Table 5.
Descriptive Statistics Per Interviewer for Items One to Three

						95% Confidence Interval for Mean	
Interviewer Code	*n*	*M*	*SD*	Min	Max	Lower Bound	Upper Bound
Item one: 'Altruism'							

2	28	4.79	.73	2.33	5.67	4.51	5.07
3	27	3.09	1.10	1.83	5.17	2.65	3.52
4	25	3.9	1.1	1.83	5.83	3.46	4.37
6	28	3.74	.80	2.00	5.17	3.43	4.065
7	22	3.29	.66	1.83	4.83	3.00	3.58
9	27	4.16	.94	2.00	5.50	3.78	4.53
10	25	4.57	.80	2.67	5.50	4.24	4.90
11	31	3.89	1.08	1.83	5.50	3.49	4.29
12	28	3.60	1.23	1.33	5.17	3.12	4.07
13	29	4.54	.86	1.83	5.50	4.22	4.88
Total	270	3.97	1.08	1.33	5.83	3.84	4.09
Item two: ‘Conscientiousness’							
2	28	4.35	1.35	1.00	6.00	3.82	4.87
3	27	5.13	.62	3.83	6.00	4.88	5.37
4	25	4.62	.88	3.29	6.00	4.26	4.99
6	28	4.24	.78	3.29	6.00	3.94	4.54
7	22	4.00	.63	3.29	6.00	3.7	4.27
9	28	4.27	1.16	2.00	6.00	3.81	4.72
10	25	4.94	.95	1.86	6.00	4.55	5.33
11	31	4.84	1.02	2.86	6.00	4.47	5.21
12	28	4.69	1.27	1.71	6.00	4.20	5.18
13	29	5.23	.78	3.43	6.00	4.94	5.53
Total	271	4.6	1.03	1.00	6.00	4.52	4.77
Item three: ‘Job satisfaction’							
2	28	5.41	.65	4.00	6.00	5.16	5.67
3	27	4.96	.88	2.50	6.00	4.61	5.31
4	24	4.88	.98	2.25	6.00	4.46	5.29
6	28	4.73	.71	3.00	5.75	4.45	5.00
7	22	4.23	.74	3.00	5.75	3.90	4.55
9	27	5.21	.85	3.25	6.00	4.88	5.55
10	25	5.13	.76	3.33	6.00	4.82	5.45
11	30	5.25	1.11	1.50	6.00	4.83	5.67
12	27	5.29	.91	2.75	6.00	4.93	5.65
13	29	4.96	1.08	2.25	6.00	4.55	5.37
Total	267	5.0218	.93	1.50	6.00	4.91	5.13

Post-hoc analyses were done in order to further investigate where differences in responses between individual interviewers lie (see Table 4, Appendix A). A summary of interviewer pairwise post-hoc comparisons for all analysed items is provided in Table 6. In the case of items one and two where equal variances could not be assumed between groups, a Games-Howell nonparametric post-hoc analysis was used to compare individual differences in mean scores between interviewers. In the case of item three where equal variances were assumed, the Tukey HSD parametric post-hoc test was used to compare individual differences in mean scores per interviewer. Interviewers were flagged to be at high risk of IE bias when they significantly differ from three others on one item. An interviewer differing from three or more others on a given item indicates that it is highly likely that consistent systematic

differences found in the scores of this interviewer when compared to several others represents some form of bias associated with the responses collected by the interviewer. This is because all interviewers collected data in the same interview conditions. Cases where interviewers statistically differ between one other on a given item may represent systematic differences due to their presence in the data collection, though it is much less likely to be the case.

Table 6 summarises cases where interviewers differed from each other per item. Interviewer 7 significantly differed to three or more interviewers on all three OCB- TP subscales; interviewers 3 and 13 significantly differed to three or more interviewers on two of the three OCB-TP subscales. All other interviewers significantly differed to one other on at least one of the OCB-TP subscales.

Table 6
Summary of Difference Tests by Interviewer Gender

Interviewer number	Gender	Items with statistically significant difference**								
		Attitudinal					Referencing Interviewer		Sensitive	
		1	2	3	4	5	6	7	8	9
2	M		x*	x	x	x*	x		x*	x*
3	F	x*	x*		x	x	x	x	x*	x*
4	M		x		x	x	x	x	x	x
6	F	x	x*		x	x	x*	x	x	x*
7	M	x*	x*	x*	x*	x*	x*	x*	x*	
9	M	x	x	x	x	x	x*	x*	x*	x
10	M		x*	x	x	x	x*	x*	x*	x
11	F	x	x	x	x	x	x*	x	x*	x
12	F		x*	x	x*	x			x	x
13	F	x*	x*		x	x	x	x	x*	x*

Note. Statistically significant difference denoted by an 'x' on that item.

* Differs with three or more interviewers on this item.

** $p < .05$

The remaining two items (four and five) in the attitudinal response category were ordinal. For these, Kruskal-Wallis tests were run to investigate any significant differences in responses between interviewers. Kruskal-Wallis tests were significant for item four ($H = 19.47$, $p < .05$) and item five ($H = 21.10$, $p < .05$), indicating that the distribution of responses differed significantly between certain interviewers. Post-hoc pairwise comparisons between interviewers show that every interviewer differs from at least one other interviewer on item four and five. Interviewer seven significantly differs from three or more interviewers on both items.

4.2.2.2. Items referencing the interviewer. H_5 asserted that interviewers systematically differed on items that contain wording that refer to the interview situation. H_6

asserted that this difference would be greater than any difference observed in attitudinal items. Items six and seven (see Table 1) were ordinal; in order to test for differences between interviewers on these items, Kruskal-Wallis tests were carried out. Statistical significance on these tests would indicate that responses on these items likely systematically varied between interviewers. Kruskal-Wallis tests were significant for item six ($H = 42.18$, $p < .01$) and item seven ($H = 21.34$, $p < .05$). This indicates that when comparing the response distributions received in both situation-referencing items, interviewers achieved scores that significantly differed between one another in some way. Post-hoc pairwise comparisons for item 6 (Figure 3, Appendix A) and item 7 (Figure 4, Appendix A) for individual interviewers show that every interviewer differed from at least one other interviewer on both items, with the exception of interviewers 2 and 12. Five of the ten interviewers significantly differ from at least three other interviewers on item six, indicating that half of the interviewers who administered this item were consistently associated with results that deviated from the expected response pattern. Interviewers 7 and 9 significantly differ from three (or more) others on both situation-referencing items. This indicates that these two interviewers systematically differed from several their peers on both items (summarized by interviewer in Figures 3 and 4, Appendix A). In sum, items that referred to the interview situation elicited responses that systematically differed across several interviewers, providing support for H_5. H_6 is supported as there were more systematic differences between interviewers for items that reference the interviewer (17 out of 20 [85%] of all cases) than there were cases of significant differences for attitudinal items (27 out of 50 [54%] of all cases).

4.2.2.3. Sensitive items. H_7 asserted that sensitive items were associated with IE bias. Kruskal-Wallis tests were used to test for statistical differences in response distributions between these items as they were measured using ordinal scales. Kruskal-Wallis tests were significant for items eight ($H = 25.71$, $p < .01$) and seven ($H = 22.67$, $p < .01$). This indicates that interviewers likely differed in their responses to these items. Post-hoc pairwise comparisons between interviewers for item 8 (Figure 5, Appendix A) and item 9 (Figure 6, Appendix A) show that every interviewer differs from at least one other both items, with the exception of interviewers 7 and 12. Interviewers 2 and 3 significantly differ from three or more interviewers on both sensitive items. Seven of the ten interviewers significantly differ from three or more interviewers on item eight, indicating that the majority of interviewers achieved response distributions that results that significantly deviated from the expected response model in this item. In other terms, these results indicate that this item elicited consistent response variance by interviewer 70% if the time. H_7 that sensitive items elicited IE bias is supported.

Overall, interviewers systematically differed for across every item analysed, across several interviewers for each item. 90 cases of interviewer-per-item scores were analysed (i.e.

ten interviewers with nine items each). 79 of all cases (87.78%) were interviewers who significantly differed from at least one other on that item. 34 out of all 90 cases (37.78%) were interviewers who differed from three or more interviewers on that item, indicating a consistent systematic difference between their scores and that of other interviewers (supporting H_3). For attitudinal items, 27 out of the 50 (54%) of cases were interviewers who differed from at least one other interviewer; 15 out of these 50 (30%) of cases were interviewers who significantly differed from three or more other interviewers (supporting H_4). For items referencing the interviewer, 17 of the 20 cases (85%) were interviewers who differed from at least one other, and 8 of these 20 cases (40%) were interviewers who significantly differed from three others (supporting H_5). H_6 is supported, as items that reference the interviewer elicits response variance more consistently than attitudinal items. Sensitive item types exhibited the most cases of response variance between interviewers - 19 of the 20 cases (95%) were interviewer scores significantly differing from one other, and 11 of these 20 cases (55%) were interviewers significantly differing from three others (supporting H_7). H_8 that sensitive items elicit more bias than other items is supported, owing to the finding that interviewers varied from several other interviewers in the most cases across these items.

4.3. Findings Related to the 'Physical Appearance' Question

In order to further gauge respondent reaction to physical appearance, respondents were asked to appraise interviewers through self-report on the item, "Overall, how would you rate the physical appearance of your interviewer?". Responses (n = 180) for this item were normally distributed between 'neutral' and 'very good' appearance (see Figure 7, Appendix A). This indicates that respondents exhibited a normally distributed spread of responses on the scores provided for this item between 'neutral' and 'very good'. This finding is significant because it was expected that respondents might skew their responses due to social desirability (Campanelli & O'muircheartaigh, 1999). Table 7 shows the frequency of each appearance rating per interviewer; Interviewer 7 received no ratings of 'very good' appearance on this item, while Interviewer 9 received the most ratings of 'very good' and 'quite good' appearance on this item when compared to their peers. Further post hoc pairwise response distribution comparisons (Figure 4, Appendix A) indicate that interviewer 9 significantly differed from four of their peers on this item ($p < .05$). This indicates that interviewer 9 consistently differed from their peers on this item, likely in a positive direction given their higher spread of scores on this item. Furthermore, these pair-wise comparisons indicate that interviewer 7's scores differed from three of their peers ($p < .05$) on this item, assumed to be in a negative direction given their lower spread of scores. Taken together, these findings support that participants had an even spread of answers across interviewers on this item, and that some interviewers received different ratings to others (as tested in 4.2.3.2). Two interviewers (numbers 7 and 9) were flagged for having significantly differed from several other interviewers on this item, and their responses

were analysed in the subsequent discussion section.

Table 7
Frequencies of Interviewer Ratings on the Physical Appearance Item

Interviewer number	Scores on categories of item 7, "Overall, how would you rate the physical appearance of your interviewer?"							Total
	Very poor	Poor	Quite poor	Neutral	Quite good	Good	Very Good	
2	0	0	1	1	3	5	4	20
3	0	0	1	4	4	5	2	22
4	0	0	0	4	4	9	3	24
6	0	0	0	5	5	3	1	21
7	0	0	0	5	8	1	0	20
9	0	0	0	3	0	8	11	27
10	0	0	0	3	2	4	11	25
11	1	3	0	3	3	8	6	30
12	0	1	1	0	4	8	5	27
13	0	1	0	6	3	4	3	24
Total	1	5	3	34	36	55	46	240

Next, the significance of these results in relation to survey data collection literature and practice is discussed.

5. Discussion

5.1. Main Findings

In this study I deployed a group of interviewers to collect surveys via a door-to-door approach across several residential areas in Cape Town, and evaluated for any systematic differences between them when survey data was collected under the same interview conditions from the same sampling pool. Interviewers consistently affected shared attitudes of respondents in items related to working life. In particular, interviewers influenced responses related to life and wage satisfaction, likelihood to participate in prosocial behaviour at work, and appraisals of the interview experience. All interviewers were shown to systematically differ from at least one of their peers nine-tenths of the time across all study items, suggesting that the presence of interviewers had an effect on the way most people responded in this survey. Interviewers were further found to significantly differ from three or more of their peers one-third of the time, evidencing that some interviewers were more prone to influence responses in some cases. With interviewers collecting data from the same sampling areas, at the same time of day, wearing the same attire, of the same ethnicity, of similar age, and of similar educational and socioeconomic backgrounds, it is argued that individual interviewer characteristics brought about bias more so than any social differences. This is because, if it is assumed that when all situational variables are kept constant, any systematic difference related to interviewers are due to individual idiosyncrasies they brought to the interview situation.

This argument is supported by the finding that some interviewers were markedly more prone to eliciting bias when compared to others. In one such case (interviewer 7), an interviewer significantly differed from more than three of their peers across every survey item. This suggests that there remains some pervasive element with this interviewer that made them more prone to introducing response variance, despite there being other interviewers of the same gender and ethnicity. The bias observed in the case of this interviewer is suspected to be due to some idiosyncrasy they brought to the interview situation, for example, in the way they presented the survey or individual items. While this study did not examine methodological differences between interviewers during the interview itself, previous studies find support that it is most likely methodological idiosyncrasies in survey delivery that brought about variance between interviewers, where situational variables and observable traits are kept constant (Rasinski et al., 1999). Methodological differences might also account for why some interviewers received consistently biased results across several items, where they might have continually relied on idiosyncratic explanations when presenting the survey, thereby eliciting a consistent response pattern that differs from their peers. Because all interviewers deployed in this study are assumed to have collected data under the same interview conditions, it is assumed

that random error between several respondents in each sampling point would cancel any systematic variance between them. It is more likely, then, that bias related to methodological differences, rather than sampling zone effects, than brought about any systematic differences in scores observed in such cases.

In order to further gauge reactions to interviewer physical appearance, respondents were asked to appraise interviewers on the item, "Overall, how would you rate the physical appearance of your interviewer?". This approach elicited differences between almost all interviewers (as expected if the question is asked by the interviewer about themselves). There was an initial concern before initiating the survey procedure that responses on this item might be severely limited due to social desirability - that respondents might adjust their answers to maintain an appropriate social interaction with their interviewer so as to not offend them (O'uircheartaigh & Campaneli, 1998). However, a normally-distributed range of attitudes on this item between 'neutral appearance' and 'very good appearance' suggests that participants were not as prone to concealing their opinions as initially suspected, although they were likely to reserve dismissing interviewer appearance as 'poor' upfront.

However, statistically significant differences found between interviewers on this item suggests that certain interviewers still varied significantly compared to their peers on this item. Specifically, two male interviewers were flagged to have received systematically differing results on this item - where each of them significantly differed from six of their peers on this item, where one (interviewer 9) received the highest rating ('very good appearance') eleven times (which also constituted the majority of his scores), while the other (interviewer 7) did not received the highest rating in any instance (see Figure 4, Appendix A for pairwise comparisons on this item). These differences might support that one interviewer received potentially biased results on account that his appearance was well-received by the majority of his participants, while his peer received potentially biased results owing to the fact that his appearance was never well-received by any of his participants.

Further evidence that physical appearance might have played a role in IE bias is found in the fact that interviewer 7 is the same interviewer that differed from several of his peers across all analysed study items. In other words, this interviewer did not receive high ratings for physical appearance at any one time – while all of his peers did – and his scores differed from several of his peers on every item. This finding supports that physical appearance had some kind of influence survey responses, though this study did not make it a primary goal to quantify the extent of this effect. There is empirical support that interviewer presentation has a minor effect on bias in the interview situation, albeit in a recruitment context (Shahani, Dipboye, &

Gehrlein, 1993). This finding is supported in the present study where comparatively lower results in physical appearance in one interviewer worked in tandem with biases found across several of their scores when compared to their peers. To emphasise, this finding might be incidental given that it only occurs in one interviewer. The systematic difference associated with this interviewer's might be influenced by another variable associated with his deployment, speculated to be methodological idiosyncrasy.

Data of ethnic differences in the present study were still analysed for the purpose of making a preliminary evaluation on how IE might be influenced by ethnic differences in the dyad. These findings are to be interpreted with caution because they included data that overrepresented one ethnic group over others, though statistical conditions for evaluating group differences were met. Further data including a more ethnically diverse sample was later collected in the larger Living Wage study and would need to be reassessed. The finding that ethnic differences in the dyad did not influence IE might stand in contrast to the 'birds of a feather' effect proposed by McPherson, Smith-Lovin and Cook (2001), which proposes that people of similar geo-economic conditions are likely to hold similar attitudes. In Cape Town, ethnic differences are a proxy through which living conditions can be understood. This is owed to the fact that the Cape Town area remains one of the most economically segregated societies in the world, and this segregation persists along racial lines as demarcated by the Group Areas Act enforced during Apartheid (Bickford-Smith, 1995). Co-occurring differences along economic and ethnic lines were assumed to lead to a 'large' social distance, as represented by ethnicity, therefore eliciting a higher degree of IE bias. The finding that ethnic categories did not influence scores contrasts this assumption.

Gender differences present in the dyad were shown to influence responses to differing degrees depending on the item. The content of the item appeared to be more important than the item type when determining IE according to gender. For example, attitudinal items related to 'slacking at work' (conscientiousness) elicited more IE only when females interviewed male subjects. At the same time, IE by item type did not differ when observing gendered dyads separately. This finding suggests that interviewees engaged with an item's substantive meaning differently depending on who the interviewer was, and it was not an inherent quality of the item type that was swayed by the presence of the interviewer. In this case, it may be that male interviewees were embarrassed to admit attitudes that suggested laziness in the presence of female interviewers, rather than any gender differences exacerbating existing bias related to item type. This finding raises further questions into how traditional gender roles are elicited by item wordings, and how much these roles affect responses according to what the item discusses. It appears that some kind of gender role is elicited by item wordings that induced men to

withhold some socially undesirable responses. This finding also stands in contrast to previous literature that foregrounded the importance of gender roles when presenting sensitive items over other items (Krumpal, 2013). Here, it appears that sociocultural item interpretations might take precedence over item type in response adjustment.

Sensitive items were the most prone to inducing response variance across all interviewers, where these items underwent the most adjustment-by-interviewer when compared to other items. In the context of this study, this finding means that respondents were most likely to adjust attitudes related to income over other attitudes depending on who asked the question. Bias on sensitive items did not significantly differ between any gender dyad configurations. This stands in contrast to the expectation that bias on sensitive items would be most marked when gender differences are present in the interview (Krumpal, 2013). Furthermore, female interviewers differed from several of their peers six out of ten times, whereas males did half of the time. In previous experiments, the opposite was found where male interviewers were more likely to be associated with response adjustment for sensitive items (Rasinski et al., 1999). This finding may stand as a contrast to studies that suggest that female interviewers are more approachable in a variety of interview contexts, and therefore less likely to be associated with response concealment (Harling et al., 2016). Given that over two-thirds of interviewers differed from three or more of their peers on sensitive items, it is supported that systematic differences on these items are unrelated to any observable trait of the interviewer themselves. This is because there was an even spread of gender and no variance in ethnicity in the interviewer pool - making it more likely that another interviewer variable is eliciting a heightened response variance on sensitive items. It is speculated that interviewer idiosyncrasies is responsible for this effect, though there was no means set up to track this in the present study (discussed under ‘strengths and limitations’).

While interviewer responses differed at an item level, there was also support for systematic differences between interviewers at a construct level, specifically, for constructs related to Organisational Citizenship Behaviour (OCB). These constructs were represented by items with moderate-to-high factor loadings - overall indicating an adequate, but not ideal match between item scores and the underlying attitudes they purport to measure. Still, interviewers significantly differed from several interviewers on these constructs one-third of the time, and differed from one other interviewer on these constructs two-thirds of the time. These findings indicate that respondents might be representing underlying psychological attitudes themselves in a varying light depending on who interviewed them, beyond simply adjusting singular responses at random in the interview. It might be that respondents engage impression management tactics in order to keep a degree of consistency in the attitudes they

may share with their interviewer. Alternatively, it might be that respondents' actual attitudes are elicited to varying degrees depending on the interviewer. Though respondent's outward attitudes might vary naturally when interviewed in separate occasions, the question is whether this variance significantly differs to the rest of the population when an interviewer asks the question. This concern is underscored if interviewers are observed to maintain methodological idiosyncrasies when conducting the interview.

5.2. Strengths and Limitations

The results of this study should be interpreted in light of its strengths and limitations. The findings of this study are based on comparisons between interviewers collecting face-to-face interview data in the same sampling zones, without any non-interviewer baseline to compare their scores against. Comparing the same survey data against an interviewer-removed method, for example, Audio Computer-Assisted Self Interviewing (ACASI; Langhaug, 2011) might have provided a baseline against which to measure the relative variance introduced by interviewers. This would have been useful to estimate a precise effect following the introduction of interviewers. At present, IE is estimated by examining systematic differences in response trends between interviewers, limiting our ability to establish a baseline against which to find the 'absolute' bias introduced by human interviewers.

The results of the present study were collected with restrictions to SES variance on both the part of the interviewer group and the respondent pool. All interviewers employed were of previously unemployed status, accepting data collection work based on immediate economical concern. This condition limited my ability to examine the effect of deploying interviewers of varying SES on survey results. Respondents in this sample also, in large part, excluded middle-class workers, owing to the economic makeup of the majority of the Cape Town municipal area. This placed a similar limitation on the variance SES of respondents. Such insights into SES effects in the interview dyad might have been valuable to survey research presently deploying interviewers into various low-income communities internationally (e.g. Carr et al., 2018), because such surveys are concerned similarly populations varying in SES, interviewed by interviewers of potentially variant SES. A more rigorous analysis of the complex ties between SES in the survey interview context is likely to yield valuable insights for quantitative research in other societies where SES might impact respondent attitudes.

Certain methodological metrics could have been better tracked during the procedure of this study, for example, interviewer methodology and rejection rates. Though survey presentation methodology was presented in an initial training, it was not checked after commencing data collection; not doing so limited my ability to examine differences between

interviewer survey presentation and barred any examination into the influence of survey delivery on various item types. Examining interviewer methodology post-training might have provided critical insights into why some interviewers received consistently biased results. In addition, rejection rates by interviewer (i.e. the frequency that respondents rejected an interviewer's approach) were not tracked during data collection due to oversight. This was potentially an important metric because some interviewers might have been accepted for interview by a different sub-sample to others, potentially influencing their results. Furthermore, some interviewers might have been rejected more regularly than others, providing insight into how some participants may have responded to interviewers of varying characteristics (e.g. gender, age, ethnicity). Finally, inquiry into respondent acceptance and rejection of interviews might have provided insight to respondent traits influencing response accuracy.

It appears that asking participants to rate interviewer appearance directly might be a viable strategy to eliciting accurate opinions on interviewer appearance. In an earlier draft of the survey not included in this study, respondents were asked to rate interviewer 'attractiveness' directly at the end of the survey, since physical attractiveness has been shown to be linked to respondent appraisals (Johnson et al., 2010). However, this item was withdrawn from a final version owing to the fact that it might have elicited social desirability biases in responses and furthermore cause discomfort to the interviewers. The item was subsequently replaced with a rating of 'physical appearance' (Table 1). While the shortcoming of this approach is that 'appearance' and 'attractiveness' are qualitatively different terms, physical appearance ratings might have still provided potential insights into the way respondents received interviewers. With regards to what might have differed in interviewer appearance, interviewers were given concession to wear their own waterproof gear and headwear, though each were requested display the designated T-shirt when approaching participants. This may have affected how their physical presentation was received by participants. Due to the limited number of interviewers within which this effect was observed, a replication study with more interviewers might be better suited to more definitely find support for the effect of appearance on IE bias

Finally, there were procedural shortcomings to this study. There was an underrepresentation of participant ethnicities other than the majority of the sample, being 'Black African', who constituted almost eighty percent of the participant pool. This shortcoming might have affected the homoscedasticity of responses when comparing multiple ethnic groups in the second hypothesis. Though Levene's test for homoscedasticity was nonsignificant between ethnic groups on the analysed items, I recommend that results pertaining to ethnicity in this study are interpreted with caution. One item, 'conscientiousness', when testing for homoscedasticity, was nonsignificant at the conventional $p = .05$ level, attaining a value of $p =$

.06, essentially meeting the conventional requirements for a group comparison by a bare minimum. In other words, the variances between ethnicities on this item may well have been insufficiently homogeneous, rendering the difference tests carried out to be statistically inappropriate. A more representative sampling pool for the multi-ethnic Cape Town working population might have allowed for ethnic group analyses to be conducted with more confidence. The findings here might inform the presently ongoing Living Wage project, which is collecting data from a more representative sampling pool.

This study, in addition, was not conducted as a true experiment with a control group and isolation of confounding variables. Therefore, we could not establish causality between the presence of interviewer and the bias they were associated with. Instead, any variables found to be influencing bias are posited tentatively. These findings are posited with advocacy towards replication in other contexts and with consideration to true-control experiments, which might establish causality at the exclusion of situational variables, for example, sampling-point bias associated with interviewer-specific rejection patterns.

5.3. Practical Implications and Future Research

Since interviewers are currently deployed to collect data across several countries in the Living Wage study, estimating the general effects of deploying these interviewers is a worthwhile undertaking. More generally, however, it remains commonplace in social sciences data collection practices to make attitudinal inferences to populations based on interviewer-collected data without correcting for potential variance introduced by the presence of interviewers (Langhaug et al., 2010). The risk forgone, should IE bias not be calculated in attitudinal surveys, is that any population inferences made are likely to be based on biased data which is unrepresentative of its population's 'true' attitudes.

Our preliminary analyses of ethnic differences found that participants did not adjust their attitudes in any significant way whether their interviewer was of the same ethnic group to themselves, or of a different one. This finding was supported when testing difference on a construct level - suggesting that the attitudes themselves, beyond individual responses, were not adjusted by respondent ethnicity. This might be of interest to research investigating the sources bias in a variety of interview contexts. However, further data will need to be collected from a more representative sample in order to get a better understanding of where interview methodology research may want to direct their attention regarding ethnic differences. Furthermore, this finding might only be applicable in sampling areas of a highly similar SES, which was the case in the present study. Different response trends may have been observed in populations where there was more diversity in SES backgrounds.

Different gender configurations in the interview influenced results in a significant way. This highlights that gender differences might be a more reliable predictor of bias in the interview situation than ethnicity; this could inform the selection of interviewers for data collection in similar studies going forward. Specifically, it was found that different-gender dyads were prone to receiving significantly biased responses for attitudes. This finding underscores the importance of selecting an even balance of interviewer gender when inferences of attitudinal scores are being made from gender-diverse samples to a gender-diverse population.

The finding that respondents adjusted their psychological attitudes consistently for almost every interviewer warrants further inquiry into the nature of the biasing effect introduced via the interviewer. I speculate that differences in attitudinal scores by interviewer go beyond social desirability in responses, and might instead be directly elicited by the interviewer via cue utilization. Realistic Accuracy Modelling (RAM; Funder, 1995) supports that attitude elicitation is a possibility when participants are responding to subtle cues in the interview. In other words, it might be the case that attitudes measured via interviews are inherently adjusted by the very presence of the interviewer, overshadowing the biasing effect brought about by any interviewer idiosyncrasy. One way to ascertain this in research would be through comparison to an objective measure (for example, biometric data), in order to better isolate the effects of the presence of an interviewer while bypassing impression management by the respondent. Such data might take quantitative survey data collection in the direction of accurately isolating – and correcting for – the effects of deploying interviewers when measuring psychological attitudes.

Though it may be a concern that without a non-interviewer group, the OCB-TP instrument's validity was likely based on IE biased data. However, the factor structure uncovered might still be useful to understand how the proposed constructs are represented in the population. Here, it is assumed that the population would respond with the same bias to the same interviewers over a sufficient sample size, cancelling out any systematic effect of response variance on factor loadings. Comparison of levels of IE to an underlying factor structure might provide valuable insight into the functioning of bias at the construct-level, and how impression management might influence 'true' attitudes beyond individual item scores.

The social implications of gender differences in responding to certain items was not a core interest to this study. Though it was interesting to observe that only males adjusted their responses for certain items in the presence of female interviewers, these results were interpreted in lieu of being one interview dyad configuration in survey methodological research. However,

this finding potentially opens up a vein of questions related to social roles being elicited by certain item wordings. Another potentially large body of research that might take interest in this finding are those that evaluate power disparities present in interview contexts, for example, qualitative research analyses by Sallee and Harris III (2011).

Overall, a more diverse interviewer group could have been sourced for the purposes of data collection. The present study included interviewers from the same youth group. While this controlled for variation in interviewer education and SES, sourcing interviewers from a more diverse background might have allowed for more analyses into the effects of interviewer SES on attitudinal data. Especially in an environment like Cape Town where SES differences have a history of being politically-charged and impactful on social interactions, it is a loss to not include SES in the present analyses on IE. The Living Wage project and similar survey methodology studies going forward may wish to be cognisant of SES variations in the interview dyad and their possible biasing consequences.

Finally, data collectors are recommended to treat methodological idiosyncrasies between interviewers with particular sensitivity, as they are suspected to play a key role in eliciting bias from respondents. This is supported by the finding that there were systematic differences between interviewers despite attempts to control for interviewer appearance, ethnicity and interviewer gender. The presence of systematic differences despite controlling for these characteristics indicates that bias is likely to persist regardless of interviewer characteristics. It is recommended that future research employ an observational method to track interviewer idiosyncrasies during data collection to further examine why individual differences between interviewers come to be. Alternatively, future research could focus attention on the effect of perceived interviewer attractiveness on response bias, as attractiveness has been shown to significantly impact interviewer ratings in recruitment contexts (Johnson et al., 2010), but is yet to be investigated in household survey contexts.

Conclusions

This study investigated whether interviewers systematically differed in a quantitative survey of workers in Cape Town. We found the most variability between interviewers when they presented sensitive questions. We found further support that most interviewers are likely to exert a biasing effect on attitudes measured across a variety of item types, including attitudinal and interview situation-referencing items. We also found support that gender differences between interviewers and interviewees systematically influenced responses. This work highlights the importance of evaluating the extent to which survey responses vary by interviewer, especially for questions related to satisfaction with wages and income. It also highlights the importance of considering gender differences when deploying interviewers to deliver surveys. It is recommended that future quantitative surveys estimate and adjust for the biasing effects of introducing an interviewer to collect data.

IE is a problem because it is likely to lead to the misinterpretation of results, where attitudinal data is taken to represent a 'true' variance of respondent attitudes with oversight to systematic error present when interviewers are deployed. This oversight is intrinsic to survey practice, as Suchman (1962) remarks, "it would appear from this long list [of antecedents] that the problem of bias is inherent in research upon human individuals" (p. 102). I agree that bias is inherent when any self-report attitude is quantified, more so when respondents are required to represent these attitudes face-to-face to a stranger. However, I advocate that the investigation of IE not be dismissed owing to the fact that awareness of its antecedents and consequences may lead to more representative data. At the same time, it may be cost-effective and closer to objectivity to remove interviewers from surveys where possible. Survey methods (for example, ACASI, Langhaug et al., 2011) have been developed to address this issue, though in practice data collectors still resort to face-to-face data collection without due consideration of the biases this method is laden with.

Analysing for biases present in any population and correcting for them systematically could uncover critical information about the systematic biases present in the population. Such population differences could be meaningful in some cases, for example, in investigating response concealment where interview conditions (e.g. class and gender differences) affect the responses collected. In sum, IE is important to track when collecting survey data, owing to its tendency to lead to data misrepresentation. Survey researchers are encouraged to take corrective action with awareness to the salience of IE in order to track the systematic biases present in survey populations.

References

Anderson, B. A., & Silver, B. D. (1986). Measurement and mismeasurement of the validity of the self-reported vote. *American Journal of Political Science*, *30*(4), 771-785.

Adida, C. L., Ferree, K. E., Posner, D. N., & Robinson, A. L. (2016). Who's asking? Interviewer coethnicity effects in African survey data. *Comparative Political Studies, 49*(12), 1630-1660.

Akaike, H. (1987). Factor analysis and AIC. In *Selected papers of Hirotogu Akaike* (pp. 371-386). Springer, New York, NY.

Bickford-Smith, V. (1995). South African urban history, racial segregation and the unique case of Cape Town? *Journal of Southern African Studies*, *21*(1), 63-78.

Campanelli, P., & O'muircheartaigh, C. (1999). Interviewers, interviewer continuity, and panel survey nonresponse. *Quality and Quantity*, *33*(1), 59-76.

Carlson, E. (2015). Ethnic voting and accountability in Africa: A choice experiment in Uganda. *World Politics*, *67*(2), 353-385.

Carr, S. C., Maleka, M., Meyer, I., Barry, M. L., Haar, J., Parker, J., ... & Young-Hausner, A. (2018). How can wages sustain a living? By getting ahead of the curve. *Sustainability Science*, *13*(4), 901-917.

Davis, R. E., Couper, M. P., Janz, N. K., Caldwell, C. H., & Resnicow, K. (2010). Interviewer effects in public health surveys. *Health Education Research, 25*(1), 14-26.

Dykema, J., Diloreto, K., Price, J. L., White, E., & Schaeffer, N. C. (2012). ACASI gender-of-interviewer voice effects on reports to questions about sensitive behaviors among young adults. *Public Opinion Quarterly*, *76*(2), 311-325.

Ellemers, N. (1993). The influence of socio-structural variables on identity management strategies. *European Review of Social Psychology, 4*(1), 27-57.

Ellemers, N., De Gilder, D., & Haslam, S. A. (2004). Motivating individuals and groups at work: A social identity perspective on leadership and group performance. *Academy of Management Review*, *29*(3), 459-478.

Ellorenco, S., Teng-Calleja, M., Bertulfo, D. J., Clemente, J. A., & Menguito, M. L. (2019). Work-Nonwork Spillover of Wage Justice through Work Capabilities in Low and Middle Income Workers. *Journal of Human Development and Capabilities*, *20*(3), 329-344.

Field, H. S., & Armenakis, A. A. (1974). On use of multiple tests of significance in psychological research. *Psychological Reports*, *35*(1), 427-431.

Funder, D. C. (1995). On the accuracy of personality judgment: a realistic approach. *Psychological review*, *102*(4), 652.

Groves, R. M., & Magilavy, L. J. (1986). Measuring and explaining interviewer effects in centralized telephone surveys. *Public Opinion Quarterly*, *50*(2), 251-266.

Harling, G., Chanda, M. M., Ortblad, K. F., Mwale, M., Chongo, S., Kanchele, C., ... & Oldenburg, C. E. (2019). The influence of interviewers on survey responses among female sex workers in Zambia. *BMC medical research methodology*, *19*(1), 1-12.

Johnson, S. K., Podratz, K. E., Dipboye, R. L., & Gibbons, E. (2010). Physical attractiveness biases in ratings of employment suitability: Tracking down the "beauty is beastly" effect. *The Journal of social psychology*, *150*(3), 301-318.

Kish, L. (1962). Studies of interviewer variance for attitudinal variables. *Journal of the American statistical association*, *57*(297), 92-115.

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity*, *47*(4), 2025-2047.

Langhaug, L. F., Cheung, Y. B., Pascoe, S. J., Chirawu, P., Woelk, G., Hayes, R. J., & Cowan, F. M. (2011). How you ask really matters: randomised comparison of four sexual behaviour questionnaire delivery modes in Zimbabwean youth. *Sexually Transmitted Infections*, *87*(2), 165-173.

Lix, L., Keselman, J., & Keselman, H. (1996). Consequences of Assumption Violations Revisited: A Quantitative Review of Alternatives to the One-Way Analysis of Variance "F" Test. *Review of Educational Research*, *66*(4), 579-619. Retrieved October 7, 2020, from http://www.jstor.org/stable/1170654

Mangione, T. W., Fowler, F. J., & Louis, T. A. (1992). Question characteristics and interviewer effects. *Journal of Official Statistics Stockholm*, *8*, 293-293.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, *27*(1), 415-444.

Municipal Demarcation Board (2016). *Municipal boundary demarcation process. Retrieved from* http://www.demarcation.org.za/site/western-cape/

O'Muircheartaigh, C., & Campanelli, P. (1998). The relative impact of interviewer effects and sample design effects on survey precision. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *161*(1), 63-77.

Pollner, M. (1998). The effects of interviewer gender in mental health interviews. *The Journal of nervous and mental disease*, *186*(6), 369-373.

Rapkin, B. D., & Luke, D. A. (1993). Cluster analysis in community research: Epistemology and practice. *American Journal of Community Psychology*, *21*(2), 247-277.

Rasinski, K. A., Willis, G. B., Baldwin, A. K., Yeh, W., & Lee, L. (1999). Methods of data collection, perceptions of risks and losses, and motivation to give truthful answers to sensitive survey questions. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *13*(5), 465-484.

Sallee, M. W., & Harris III, F. (2011). Gender performance in qualitative studies of masculinities. Qualitative research, 11(4), 409-429.

Schnell, R. and Kreuter, F. (2005). Separating Interviewer and Sampling-Point Effects. *Journal of Official Statistics, 21,* 389-410

Shahani, C., Dipboye, R. L., & Gehrlein, T. M. (1993). Attractiveness bias in the interview: Exploring the boundaries of an effect. *Basic and Applied Social Psychology*, *14*(3), 317-328.

Shao, J. (1997). An asymptotic theory for linear model selection. *Statistica sinica*, *7*(1). 221-242.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359-1366.

Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of personality assessment*, *80*(1), 99-103.

Stocké, V. (2004). The interdependence of determinants for the strength and direction of social disirability bias in racial attitude surveys. *Sonderforschungbereich*, *4*(14), 1-27

Stocké, V. (2007). Determinants and consequences of survey respondents' social desirability beliefs about racial attitudes. *Methodology*, *3*(3), 125-138.

Suchman, E. A. (1962). An analysis of "bias" in survey research. *Public Opinion Quarterly*, *26*(1) 102-111.

Tourangeau, R., Rasinski, K., Jobe, J. B., Smith, T. W., & Pratt, W. F. (1997). Sources of error in a survey on sexual behaviour. *Journal Of Official Statistics Stockholm*, *13*, 341-366.

Tu S-H, Liao P-S. (2007). Social distance, respondent cooperation and item nonresponse in sex survey. *Quality and Quantity, 41*(2) 177–99.

West B. T., Blom A. G. (2017). Explaining interviewer effects: a research synthesis. *Journal of Statistical Methods, 5*, 175–211.

Williams, L. J., & Anderson, S. E. (1991). Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. *Journal of Management*, *17*(3), 601-617.

Yong, A. G., & Pearce, S. (2013). A beginner’s guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in quantitative methods for psychology*, *9*(2), 79-94.

Appendix A: Supplementary figures and tables

Table 1
First Iteration of Factor Analysis of the OCB-TP

Item number	Factor				
	1	2	3	4	5
Eigenvalues	5.27	4.65	2.20	1.27	1.02
% Explained variance	23.96	21.17	9.99	5.77	4.67
OCB-TP item 1					-.785
OCB-TP item 2					-.697
OCB-TP item 3					-.635
OCB-TP item 4	.532				
OCB-TP item 5	.791				
OCB-TP item 7	.467				
OCB-TP item 16			-.511	.362	
OCB-TP item 19			-.598		
OCB-TP item 20			-.755		
OCB-TP item 21			-.851		
OCB-TP item 22			-.753		
OCB-TP item 10 (R)		.739			
OCB-TP item 11 (R)		.815			
OCB-TP item 12 (R)		.730			
OCB-TP item 14 (R)		.809			
OCB-TP item 15 (R)		.760			
OCB-TP item 18 (R)		.643			
OCB-TP item 13		-.654			
OCB-TP item 4	.465				-.387
OCB-TP item 8	.205			.284	
OCB-TP item 9				.640	
OCB-TP item 17			-.605		

Note. N = 282. The extraction method was principal axis factoring. Rotation Method: Oblimin with Kaiser Normalization. Factor loadings below .30 removed. Reverse-scored items denoted with an (R).

Table 2
Second iteration of OCB-TP EFA with factor loadings below .30 and cross-loadings removed

Item number		Factor			
		1	2	3	4
	Eigenvalues	4.63	4.16	1.97	1.07
	% Explained variance	25.75	23.16	10.94	5.93
OCB-TP item 1				.662	
OCB-TP item 2				.806	
OCB-TP item 3				.718	
OCB-TP item 5				.625	
OCB-TP item 6				.644	
OCB-TP item 7				.496	
OCB-TP item 16		.368			.473
OCB-TP item 19		.521			
OCB-TP item 20		.815			
OCB-TP item 21		.899			
OCB-TP item 22		.685			
OCB-TP item 10			.718		
OCB-TP item 11 (R)			.805		
OCB-TP item 12 (R)			.697		
OCB-TP item 14 (R)			.845		
OCB-TP item 15 (R)			.692		
OCB-TP item 18 (R)			.816		
OCB-TP item 17		.441			.509

Note. N = 282. Items 4, 8, 9 and 13 removed from first iteration denoted by Table 1. Rotation Method: Oblimin with Kaiser Normalization. Factor loadings below .30 removed. Reverse-scored items denoted with an (R).

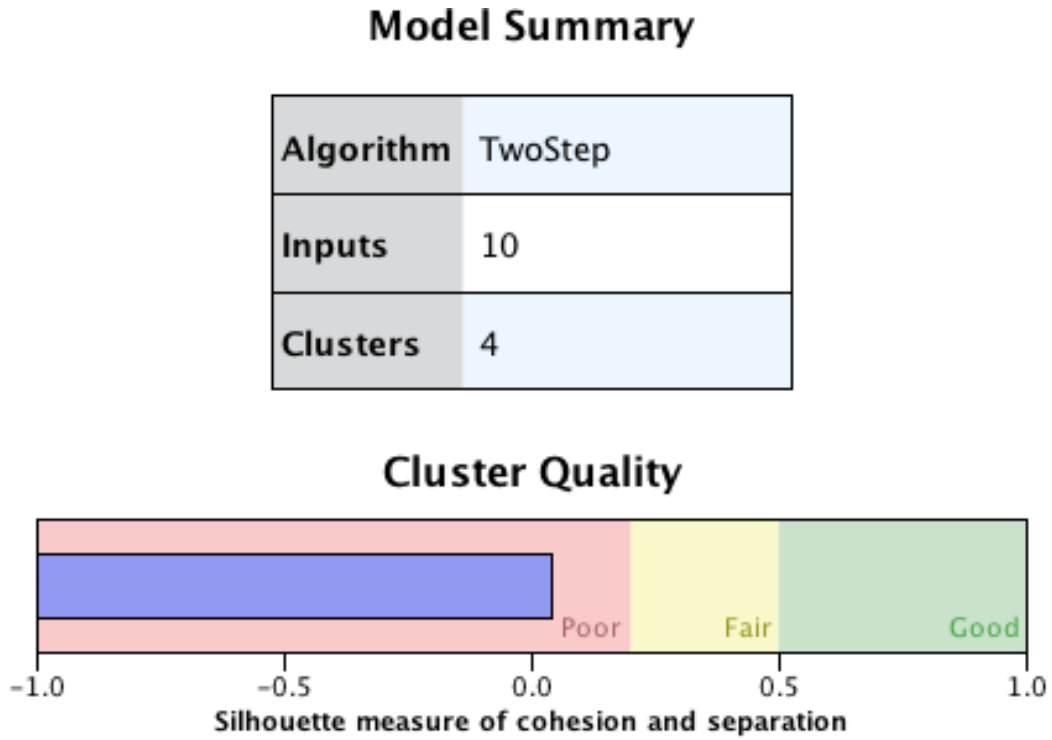

Figure 1. Two-step cluster model summary of all study items (*n* = 9).

Figure 2. Two-step cluster model summary of items in the attitudinal category (*n* = 5).

Table 3
Post Hoc Pairwise Comparisons of Gender Dyads scores on the OCB-TP

Item	(I) Gender Dyad	(J) Gender Dyad	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval of mean	
						Lower Bound	Upper Bound
Altruism	1	2	-.04	.20	.86	-.43	.36
		3	-.42*	.20	.03	-.82	-.03
		4	-.43*	.19	.03	-.81	-.04
	2	1	.04	.20	.86	-.36	.43
		3	-.39*	.19	.04	-.76	-.02
		4	-.39*	.18	.04	-.75	-.03
	3	1	.43*	.20	.03	.03	.82
		2	.39*	.19	.04	.02	.76
		4	.00	.18	1.00	-.36	.36
	4	1	.43*	.19	.03	.04	.81
		2	.39*	.18	.04	.03	.75
		3	.00	.18	1.00	-.36	.36
Conscientiousness	1	2	-.21	.21	.34	-.63	.22
		3	.47*	.21	.03	.05	.89
		4	.13	.21	.52	-.27	.54
	2	1	.21	.21	.34	-.22	.63
		3	.68*	.20	.00	.28	1.07
		4	.34	.20	.09	-.05	.72
	3	1	-.47*	.21	.03	-.89	-.05
		2	-.68*	.20	.00	-1.07	-.28
		4	-.34	.19	.08	-.72	.04
	4	1	-.13	.21	.52	-.54	.27
		2	-.34	.20	.09	-.72	.05
		3	.34	.19	.08	-.04	.72
Job Satisfaction	1	2	.18	.19	.34	-.19	.56
		3	-.16	.19	.41	-.53	.22
		4	.23	.18	.21	-.13	.59
	2	1	-.18	.19	.34	-.56	.19
		3	-.34	.18	.06	-.69	.02
		4	.05	.17	.80	-.30	.39
	3	1	.16	.19	.41	-.22	.53
		2	.34	.18	.06	-.02	.69

	4	.38*	.17	.03	.04	.72
4	1	-.23	.18	.21	-.59	.13
	2	-.05	.17	.80	-.39	.30
	3	-.38*	.17	.03	-.72	-.04

Note. *The mean difference is significant at the .05 level.

Table 4
Post Hoc Pairwise Comparisons by Interviewer for the OCB-TP

Item (test)	(I) Int. Number	(J) Int Number	Mean Dif. (I-J)	SD	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Altruism (Games-Howell)	2	3	-.78	.28	.19	-1.72	.17
		4	-.27	.31	1.00	-1.29	.76
		6	.12	.29	1.00	-.86	1.09
		7	.36	.29	.96	-.61	1.32
		9	.08	.34	1.00	-1.03	1.19
		10	-.59	.32	.70	-1.64	.47
		11	-.49	.31	.86	-1.53	.55
		12	-.34	.35	.99	-1.49	.82
		13	-.88	.29	.11	-1.86	.09
	3	2	.78	.28	.19	-.17	1.72
		4	.51	.21	.35	-.20	1.22
		6	.89*	.19	.00	.27	1.52
		7	1.13*	.18	.00	.54	1.73
		9	.86*	.25	.04	.02	1.69
		10	.19	.22	1.00	-.56	.94
		11	.29	.22	.95	-.44	1.01
		12	.44	.27	.82	-.46	1.33
		13	-.11	.19	1.00	-.72	.51
	4	2	.27	.31	1.00	-.76	1.29
		3	-.51	.21	.35	-1.22	.20
		6	.38	.23	.80	-.38	1.14
		7	.62	.22	.16	-.11	1.36
		9	.35	.28	.96	-.58	1.28
		10	-.32	.26	.96	-1.17	.54
		11	-.22	.25	1.00	-1.06	.62
		12	-.07	.30	1.00	-1.05	.92
		13	-.61	.23	.20	-1.37	.14
	6	2	-.12	.29	1.00	-1.09	.86
		3	-.89*	.19	.00	-1.52	-.27
		4	-.38	.23	.80	-1.14	.38
		7	.24	.20	.97	-.42	.90
		9	-.03	.26	1.00	-.91	.84
		10	-.70	.24	.13	-1.50	.10
		11	-.61	.24	.26	-1.38	.17
		12	-.45	.28	.84	-1.39	.48

	13	-1.00*	.21	.00	-1.68	-.32
7	2	-.36	.29	.96	-1.32	.61
	3	-1.13*	.18	.00	-1.73	-.54
	4	-.62	.22	.16	-1.36	.11
	6	-.24	.20	.97	-.90	.42
	9	-.27	.26	.99	-1.13	.58
	10	-.94*	.23	.01	-1.72	-.17
	11	-.85*	.23	.02	-1.60	-.09
	12	-.69	.27	.28	-1.61	.22
	13	-1.23*	.20	.00	-1.89	-.59
9	2	-.08	.34	1.00	-1.19	1.03
	3	-.86*	.25	.04	-1.69	-.02
	4	-.35	.28	.96	-1.28	.58
	6	.03	.26	1.00	-.84	.91
	7	.27	.26	.99	-.58	1.13
	10	-.67	.29	.41	-1.63	.29
	11	-.57	.29	.61	-1.52	.37
	12	-.42	.32	.95	-1.49	.65
	13	-.96*	.26	.02	-1.84	-.09
10	2	.59	.32	.70	-.47	1.64
	3	-.19	.22	1.00	-.94	.56
	4	.32	.26	.96	-.54	1.17
	6	.70	.24	.13	-.10	1.50
	7	.94*	.23	.01	.17	1.72
	9	.67	.29	.41	-.29	1.63
	11	.10	.26	1.00	-.78	.97
	12	.25	.31	1.00	-.76	1.26
	13	-.30	.24	.96	-1.09	.50
11	2	.49	.31	.86	-.55	1.53
	3	-.29	.22	.95	-1.01	.44
	4	.22	.25	1.00	-.62	1.06
	6	.61	.24	.26	-.17	1.38
	7	.85*	.23	.02	.09	1.60
	9	.57	.29	.61	-.37	1.52
	10	-.10	.26	1.00	-.97	.78
	12	.15	.30	1.00	-.84	1.15
	13	-.39	.23	.81	-1.16	.38
12	2	.34	.35	.99	-.82	1.49
	3	-.44	.27	.82	-1.33	.46
	4	.07	.30	1.00	-.92	1.05
	6	.45	.28	.84	-.48	1.39

		7	.69	.27	.28	-.22	1.61
		9	.42	.32	.95	-.65	1.49
		10	-.25	.31	1.00	-1.26	.76
		11	-.15	.30	1.00	-1.15	.84
		13	-.55	.28	.64	-1.48	.39
	13	2	.88	.29	.11	-.09	1.86
		3	.11	.19	1.00	-.51	.72
		4	.61	.23	.20	-.14	1.37
		6	.99*	.21	.00	.32	1.68
		7	1.24*	.20	.00	.59	1.89
		9	.96*	.26	.02	.09	1.84
		10	.30	.24	.96	-.50	1.09
		11	.39	.23	.81	-.38	1.16
		12	.55	.28	.64	-.39	1.48
Conscientiou sness	2	3	1.71*	.25	.00	.86	2.55
		4	.88*	.26	.05	.01	1.75
		6	1.05*	.20	.00	.38	1.73
		7	1.50*	.20	.00	.85	2.16
		9	.63	.23	.17	-.12	1.39
		10	.22	.21	.99	-.48	.92
		11	.90*	.24	.01	.11	1.69
		12	1.19*	.27	.00	.30	2.09
		13	.25	.21	.97	-.44	.94
	3	2	-1.71*	.25	.00	-2.55	-.86
		4	-.82	.31	.21	-1.84	.19
		6	-.65	.26	.29	-1.52	.21
		7	-.20	.26	1.00	-1.05	.65
		9	-1.07*	.28	.01	-2.00	-.15
		10	-1.48*	.27	.00	-2.37	-.60
		11	-.80	.29	.17	-1.76	.15
		12	-.51	.31	.83	-1.55	.53
		13	-1.46*	.27	.00	-2.34	-.58
	4	2	-.88*	.26	.05	-1.75	-.01
		3	.82	.31	.21	-.19	1.84
		6	.17	.27	1.00	-.72	1.06
		7	.62	.26	.37	-.26	1.50
		9	-.25	.29	1.00	-1.20	.70
		10	-.66	.27	.34	-1.57	.25
		11	.02	.29	1.00	-.95	1.00
		12	.31	.32	.99	-.74	1.37

	13	-.63	.27	.40	-1.54	.27
6	2	-1.05*	.20	.00	-1.73	-.38
	3	.65	.26	.29	-.21	1.52
	4	-.17	.27	1.00	-1.06	.72
	7	.45	.21	.48	-.23	1.14
	9	-.42	.24	.75	-1.20	.36
	10	-.83*	.22	.01	-1.56	-.10
	11	-.15	.25	1.00	-.97	.66
	12	.14	.28	1.00	-.78	1.06
	13	-.80*	.22	.02	-1.52	-.08
7	2	-1.50*	.20	.00	-2.16	-.85
	3	.20	.26	1.00	-.65	1.05
	4	-.62	.26	.37	-1.50	.26
	6	-.45	.21	.48	-1.14	.23
	9	-.87*	.23	.02	-1.64	-.10
	10	-1.28*	.21	.00	-1.99	-.57
	11	-.60	.24	.30	-1.40	.20
	12	-.31	.27	.98	-1.21	.60
	13	-1.25*	.21	.00	-1.96	-.55
9	2	-.63	.23	.17	-1.39	.12
	3	1.07*	.28	.01	.15	2.00
	4	.25	.29	1.00	-.70	1.20
	6	.42	.24	.75	-.36	1.20
	7	.87*	.23	.02	.10	1.64
	10	-.41	.24	.79	-1.21	.39
	11	.27	.27	.99	-.61	1.15
	12	.56	.29	.67	-.41	1.54
	13	-.38	.24	.85	-1.18	.41
10	2	-.22	.21	.99	-.92	.48
	3	1.48*	.27	.00	.60	2.37
	4	.66	.27	.34	-.25	1.57
	6	.83*	.22	.01	.10	1.56
	7	1.28*	.21	.00	.57	1.99
	9	.41	.24	.79	-.39	1.21
	11	.68	.25	.21	-.15	1.51
	12	.97*	.28	.04	.04	1.91
	13	.03	.23	1.00	-.72	.77
11	2	-.90*	.24	.01	-1.69	-.11
	3	.80	.29	.17	-.15	1.76
	4	-.02	.29	1.00	-1.00	.95

		6	.15	.25	1.00	-.66	.97
		7	.60	.24	.30	-.20	1.40
		9	-.27	.27	.99	-1.15	.61
		10	-.68	.25	.21	-1.51	.15
		12	.29	.30	.99	-.71	1.29
		13	-.65	.25	.25	-1.48	.18
	12	2	-1.20*	.27	.00	-2.09	-.30
		3	.51	.31	.83	-.53	1.55
		4	-.31	.32	.99	-1.37	.74
		6	-.14	.28	1.00	-1.06	.78
		7	.31	.27	.98	-.60	1.21
		9	-.56	.29	.67	-1.54	.41
		10	-.97*	.28	.04	-1.91	-.04
		11	-.29	.30	.99	-1.29	.71
		13	-.94*	.28	.04	-1.88	-.01
	13	2	-.25	.21	.97	-.94	.44
		3	1.46*	.27	.00	.58	2.34
		4	.63	.27	.40	-.27	1.54
		6	.80*	.22	.02	.08	1.52
		7	1.25*	.21	.00	.55	1.96
		9	.38	.24	.85	-.41	1.18
		10	-.03	.23	1.00	-.77	.72
		11	.65	.25	.25	-.18	1.48
		12	.95*	.28	.04	.01	1.88
Job Satisfaction (HSD)	2	3	.45	.24	.67	-.31	1.22
		4	.53	.25	.49	-.25	1.32
		6	.68	.24	.12	-.07	1.44
		7	1.18*	.25	.00	.38	1.99
		9	.20	.24	1.00	-.56	.96
		10	.28	.24	.98	-.50	1.06
		11	.16	.23	1.00	-.58	.90
		12	.12	.24	1.00	-.64	.89
		13	.45	.23	.65	-.30	1.20
	3	2	-.45	.24	.67	-1.22	.31
		4	.08	.25	1.00	-.72	.87
		6	.23	.24	.99	-.54	.99
		7	.73	.25	.12	-.08	1.54
		9	-.25	.24	.99	-1.02	.52
		10	-.18	.25	1.00	-.96	.61
		11	-.29	.24	.96	-1.04	.46
		12	-.33	.24	.94	-1.10	.44

	13	.00	.24	1.00	-.76	.76
4	2	-.53	.25	.49	-1.32	.25
	3	-.08	.25	1.00	-.87	.72
	6	.15	.25	1.00	-.64	.94
	7	.65	.26	.28	-.18	1.49
	9	-.33	.25	.95	-1.13	.46
	10	-.25	.25	.99	-1.06	.55
	11	-.37	.24	.88	-1.15	.40
	12	-.41	.25	.83	-1.20	.39
	13	-.08	.24	1.00	-.86	.70
6	2	-.68	.24	.12	-1.44	.07
	3	-.23	.24	.99	-.99	.54
	4	-.15	.25	1.00	-.94	.64
	7	.50	.25	.61	-.30	1.31
	9	-.48	.24	.59	-1.24	.28
	10	-.40	.24	.82	-1.18	.37
	11	-.52	.23	.44	-1.26	.22
	12	-.56	.24	.37	-1.32	.21
	13	-.23	.23	.99	-.98	.52
7	2	-1.18*	.25	.00	-1.99	-.38
	3	-.73	.25	.12	-1.54	.08
	4	-.65	.26	.28	-1.49	.18
	6	-.50	.25	.61	-1.31	.30
	9	-.98*	.25	.01	-1.80	-.17
	10	-.91*	.26	.02	-1.73	-.08
	11	-1.02*	.25	.00	-1.82	-.23
	12	-1.06*	.25	.00	-1.87	-.25
	13	-.73	.25	.11	-1.53	.07
9	2	-.20	.24	1.00	-.96	.56
	3	.25	.24	.99	-.52	1.02
	4	.33	.25	.95	-.46	1.13
	6	.48	.24	.59	-.28	1.24
	7	.98*	.25	.01	.17	1.80
	10	.08	.25	1.00	-.71	.86
	11	-.04	.24	1.00	-.79	.71
	12	-.08	.24	1.00	-.85	.69
	13	.25	.24	.99	-.50	1.01
10	2	-.28	.24	.98	-1.06	.50
	3	.18	.25	1.00	-.61	.96
	4	.25	.25	.99	-.55	1.06
	6	.40	.24	.82	-.37	1.18

	7	.91*	.26	.02	.08	1.73
	9	-.08	.25	1.00	-.86	.71
	11	-.12	.24	1.00	-.88	.65
	12	-.15	.25	1.00	-.94	.63
	13	.18	.24	1.00	-.60	.95
11	2	-.16	.23	1.00	-.90	.58
	3	.29	.24	.96	-.46	1.04
	4	.37	.24	.88	-.40	1.15
	6	.52	.23	.44	-.22	1.26
	7	1.02*	.25	.00	.23	1.82
	9	.04	.24	1.00	-.71	.79
	10	.12	.24	1.00	-.65	.88
	12	-.04	.24	1.00	-.79	.71
	13	.29	.23	.96	-.44	1.03
12	2	-.12	.24	1.00	-.89	.64
	3	.33	.24	.94	-.44	1.10
	4	.41	.25	.83	-.39	1.20
	6	.56	.24	.37	-.21	1.32
	7	1.06*	.25	.00	.25	1.87
	9	.08	.24	1.00	-.69	.85
	10	.15	.25	1.00	-.63	.94
	11	.04	.24	1.00	-.71	.79
	13	.33	.24	.93	-.43	1.09
13	2	-.45	.23	.65	-1.20	.30
	3	.00	.24	1.00	-.76	.76
	4	.08	.24	1.00	-.70	.86
	6	.23	.23	.99	-.52	.98
	7	.73	.25	.11	-.07	1.53
	9	-.25	.24	.99	-1.01	.50
	10	-.18	.24	1.00	-.95	.60
	11	-.29	.23	.96	-1.03	.44
	12	-.33	.24	.93	-1.09	.43

Note. *The mean difference is significant at a .05 level

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
7-9	-94.069	19.372	-4.856	.000	.000
7-10	-102.523	19.708	-5.202	.000	.000
7-12	-76.180	19.372	-3.932	.000	.004
7-4	64.963	19.708	3.296	.001	.044
13-10	64.963	19.027	3.414	.001	.029
11-10	55.828	18.171	3.072	.002	.096
6-10	-61.118	19.708	-3.101	.002	.087
13-9	56.509	18.679	3.025	.002	.112
3-10	-54.357	19.463	-2.793	.005	.235
6-9	-52.664	19.372	-2.719	.007	.295
11-9	47.374	17.806	2.661	.008	.351
7-2	58.112	22.094	2.630	.009	.384
7-11	-46.695	19.078	-2.448	.014	.647
3-9	-45.903	19.123	-2.400	.016	.737
7-3	48.166	20.312	2.371	.018	.798
2-10	-44.411	21.316	-2.083	.037	1.000
13-12	38.620	18.679	2.068	.039	1.000
7-6	41.405	20.547	2.015	.044	1.000
4-10	-37.560	18.832	-1.995	.046	1.000
7-13	-37.560	19.895	-1.888	.059	1.000
6-12	-34.775	19.372	-1.795	.073	1.000
2-9	-35.957	21.006	-1.712	.087	1.000
11-12	-29.485	17.806	-1.656	.098	1.000
4-9	-29.106	18.480	-1.575	.115	1.000
3-12	-28.014	19.123	-1.465	.143	1.000
13-4	27.403	19.027	1.440	.150	1.000
12-10	26.343	18.480	1.426	.154	1.000
6-4	23.558	19.708	1.195	.232	1.000
11-4	18.268	18.171	1.005	.315	1.000
12-9	17.889	18.121	.987	.324	1.000
13-2	20.552	21.489	.956	.339	1.000
3-4	-16.797	19.463	-.863	.388	1.000
2-12	-18.068	21.006	-.860	.390	1.000
6-2	16.707	22.094	.756	.450	1.000
4-12	-11.217	18.480	-.607	.544	1.000
11-2	11.417	20.734	.551	.582	1.000
13-3	10.606	19.652	.540	.589	1.000
13-11	9.135	18.373	.497	.619	1.000
9-10	-8.454	18.480	-.457	.647	1.000
3-2	9.946	21.876	.455	.649	1.000
6-3	6.761	20.312	.333	.739	1.000
2-4	-6.851	21.316	-.321	.748	1.000
6-11	-5.290	19.078	-.277	.782	1.000
13-6	3.845	19.895	.193	.847	1.000
11-3	1.471	18.824	.078	.938	1.000

Figure 3. Post hoc comparisons of observed response distributions by interviewer using the Kruskal-Wallis test for item 6. Each row tests the null hypothesis that sample 1 and sample 2 distributions are the same. The significance level is $p < .05$.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
7-6	13.357	19.095	.700	.484	1.000
7-13	-21.162	18.233	-1.161	.246	1.000
7-3	23.375	18.488	1.264	.206	1.000
7-11	-34.333	16.990	-2.021	.043	1.000
7-4	37.125	17.604	2.109	.035	1.000
7-2	46.607	19.095	2.441	.015	.659
7-12	-47.829	17.794	-2.688	.007	.324
7-10	-65.900	17.604	-3.743	.000	.008
7-9	-69.023	17.272	-3.996	.000	.003
6-13	-7.805	18.233	-.428	.669	1.000
6-3	10.018	18.488	.542	.588	1.000
6-11	-20.976	16.990	-1.235	.217	1.000
6-4	23.768	17.604	1.350	.177	1.000
6-2	33.250	19.095	1.741	.082	1.000
6-12	-34.472	17.794	-1.937	.053	1.000
6-10	-52.543	17.604	-2.985	.003	.128
6-9	-55.666	17.272	-3.223	.001	.057
13-3	2.213	17.597	.126	.900	1.000
13-11	13.172	16.015	.822	.411	1.000
13-4	15.963	16.666	.958	.338	1.000
13-2	25.445	18.233	1.396	.163	1.000
13-12	26.667	16.866	1.581	.114	1.000
13-10	44.738	16.666	2.684	.007	.327
13-9	47.861	16.314	2.934	.003	.151
3-11	-10.958	16.305	-.672	.502	1.000
3-4	-13.750	16.945	-.811	.417	1.000
3-2	23.232	18.488	1.257	.209	1.000
3-12	-24.454	17.142	-1.427	.154	1.000
3-10	-42.525	16.945	-2.510	.012	.544
3-9	-45.648	16.599	-2.750	.006	.268
11-4	2.792	15.296	.183	.855	1.000
11-2	12.274	16.990	.722	.470	1.000
11-12	-13.496	15.514	-.870	.384	1.000
11-10	31.567	15.296	2.064	.039	1.000
11-9	34.689	14.911	2.326	.020	.900
4-2	9.482	17.604	.539	.590	1.000
4-12	-10.704	16.185	-.661	.508	1.000
4-10	-28.775	15.976	-1.801	.072	1.000
4-9	-31.898	15.608	-2.044	.041	1.000
2-12	-1.222	17.794	-.069	.945	1.000
2-10	-19.293	17.604	-1.096	.273	1.000
2-9	-22.416	17.272	-1.298	.194	1.000
12-10	18.071	16.185	1.117	.264	1.000
12-9	21.194	15.822	1.340	.180	1.000
10-9	3.123	15.608	.200	.841	1.000

Figure 4. Post hoc comparisons of observed response distributions by interviewer using the Kruskal-Wallis test for item 7. Each row tests the null hypothesis that sample 1 and sample 2 distributions are the same. The significance level is $p < .05$.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
13-3	76.697	20.589	3.725	.000	.009
9-3	69.157	21.082	3.280	.001	.047
10-3	65.674	21.270	3.088	.002	.091
13-11	54.741	18.870	2.901	.004	.167
4-3	60.475	21.688	2.788	.005	.238
12-3	57.807	20.907	2.765	.006	.256
13-2	51.161	19.038	2.687	.007	.324
9-11	-47.201	19.407	-2.432	.015	.675
7-3	55.398	22.751	2.435	.015	.670
6-3	50.176	21.082	2.380	.017	.779
9-2	43.621	19.570	2.229	.026	1.000
10-11	-43.719	19.610	-2.229	.026	1.000
10-2	40.139	19.772	2.030	.042	1.000
4-11	-38.519	20.063	-1.920	.055	1.000
12-11	35.851	19.216	1.866	.062	1.000
4-2	34.939	20.221	1.728	.084	1.000
12-2	32.271	19.381	1.665	.096	1.000
7-11	-33.443	21.208	-1.577	.115	1.000
6-11	-28.220	19.407	-1.454	.146	1.000
7-2	29.863	21.358	1.398	.162	1.000
13-6	26.521	19.407	1.367	.172	1.000
6-2	24.640	19.570	1.259	.208	1.000
2-3	-25.536	20.743	-1.231	.218	1.000
11-3	21.956	20.589	1.066	.286	1.000
13-7	21.299	21.208	1.004	.315	1.000
13-12	18.890	19.216	.983	.326	1.000
9-6	18.981	19.929	.952	.341	1.000
13-4	16.222	20.063	.809	.419	1.000
10-6	15.498	20.127	.770	.441	1.000
9-7	13.758	21.687	.634	.526	1.000
9-12	-11.350	19.744	-.575	.565	1.000
13-10	11.023	19.610	.562	.574	1.000
4-6	-10.299	20.569	-.501	.617	1.000
10-7	10.276	21.869	.470	.638	1.000
9-4	8.681	20.569	.422	.673	1.000
10-12	-7.867	19.944	-.394	.693	1.000
13-9	7.540	19.407	.389	.698	1.000
12-6	7.631	19.744	.387	.699	1.000
10-4	5.199	20.761	.250	.802	1.000
7-6	5.223	21.687	.241	.810	1.000
4-7	-5.077	22.276	-.228	.820	1.000
2-11	-3.580	19.038	-.188	.851	1.000
9-10	-3.482	20.127	-.173	.863	1.000
4-12	-2.668	20.389	-.131	.896	1.000
12-7	2.408	21.517	.112	.911	1.000

Figure 5. Post hoc comparisons of observed response distributions by interviewer using the Kruskal-Wallis test for item 8. Each row tests the null hypothesis that sample 1 and sample 2 distributions are the same. The significance level is $p < .05$.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
13-2	62.643	19.493	3.214	.001	.059
12-2	60.118	19.864	3.026	.002	.111
9-2	54.761	19.672	2.784	.005	.242
13-3	53.313	19.864	2.684	.007	.327
12-3	50.788	20.229	2.511	.012	.542
13-6	49.401	19.672	2.511	.012	.541
10-2	49.290	20.525	2.402	.016	.735
12-6	46.876	20.040	2.339	.019	.870
11-2	45.292	19.324	2.344	.019	.859
9-3	45.431	20.040	2.267	.023	1.000
9-6	41.519	19.850	2.092	.036	1.000
4-2	41.443	20.289	2.043	.041	1.000
7-2	41.622	21.678	1.920	.055	1.000
10-3	39.961	20.878	1.914	.056	1.000
11-3	35.962	19.698	1.826	.068	1.000
10-6	36.048	20.695	1.742	.082	1.000
11-6	32.050	19.505	1.643	.100	1.000
4-3	32.114	20.646	1.555	.120	1.000
7-3	32.293	22.013	1.467	.142	1.000
4-6	-28.201	20.461	-1.378	.168	1.000
7-6	28.380	21.840	1.299	.194	1.000
13-4	21.199	20.289	1.045	.296	1.000
13-7	21.021	21.678	.970	.332	1.000
12-4	18.675	20.646	.905	.366	1.000
13-11	17.351	19.324	.898	.369	1.000
12-7	18.496	22.013	.840	.401	1.000
12-11	14.826	19.698	.753	.452	1.000
6-2	13.242	19.672	.673	.501	1.000
9-4	13.317	20.461	.651	.515	1.000
13-10	13.352	20.525	.651	.515	1.000
9-7	13.138	21.840	.602	.547	1.000
12-10	10.828	20.878	.519	.604	1.000
9-11	-9.469	19.505	-.485	.627	1.000
3-2	9.330	19.864	.470	.639	1.000
13-9	7.882	19.672	.401	.689	1.000
10-4	7.847	21.282	.369	.712	1.000
10-7	7.668	22.611	.339	.735	1.000
12-9	5.358	20.040	.267	.789	1.000
9-10	-5.470	20.695	-.264	.792	1.000
10-11	-3.999	20.365	-.196	.844	1.000
6-3	3.912	20.040	.195	.845	1.000
11-4	3.848	20.127	.191	.848	1.000
11-7	3.670	21.527	.170	.865	1.000
13-12	2.525	19.864	.127	.899	1.000
7-4	.179	22.397	.008	.994	1.000

Figure 6. Post hoc comparisons of observed response distributions by interviewer using the Kruskal-Wallis test for item 9. Each row tests the null hypothesis that sample 1 and sample 2 distributions are the same. The significance level is $p < .05$.

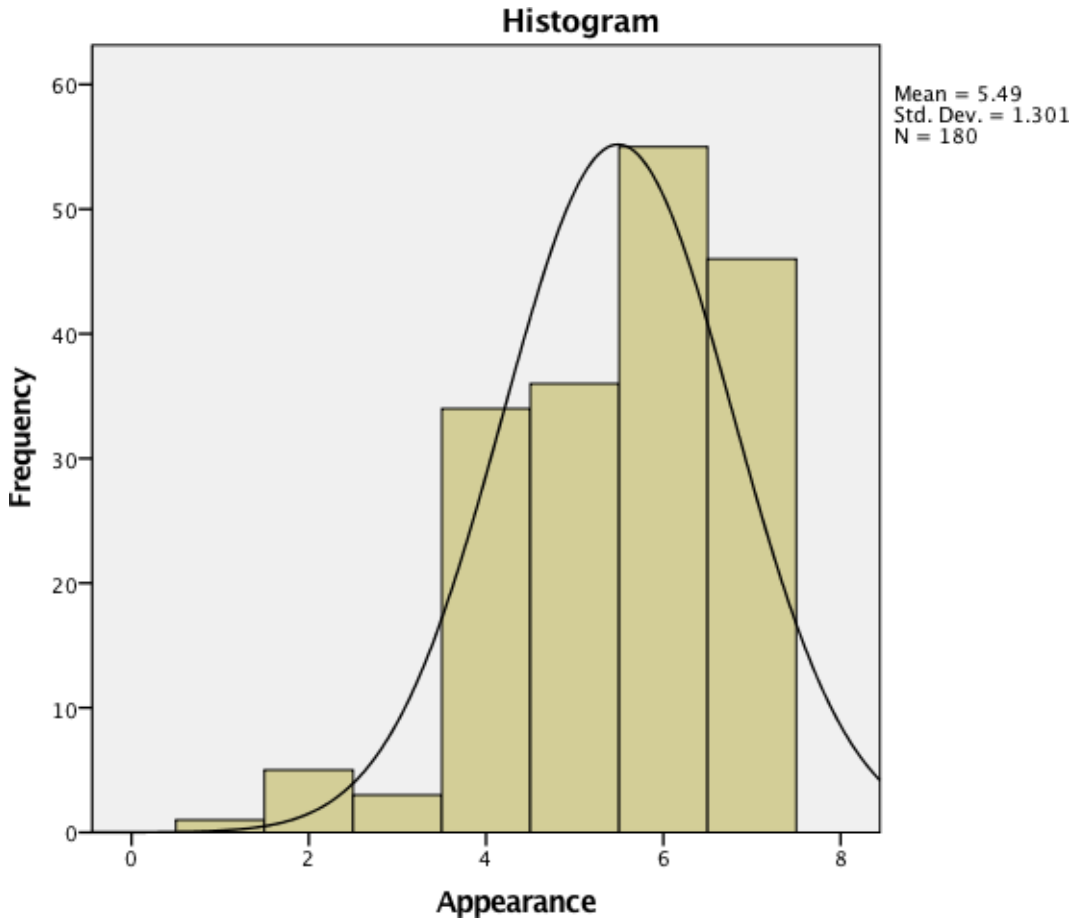

Figure 7. Histogram depicting a normal distribution of responses for item 7.

Appendix B: Living Wage Survey Instrument

SURVEY ON SOUTH AFRICANS' WELLBEING AT WORK AND IN LIFE

Respondent No._____Ward _____

Interviewer _____

Good morning/afternoon. I am_____from the University of Cape Town. We are conducting research that looks into the quality of living and quality of work of South Africans. We would like to know what South Africans consider important in having a good life and how possible it is for them to achieve this.

Yours was one of the households chosen to participate in this study. If there is someone living in your household who works and who is willing to assist us, I will be asking the person questions regarding their work, household, health, and other aspects of daily living. This should take no more than 30 minutes.

All information you share with me will be treated confidentially and will only be used for research purposes. We are not recording your name, address or phone number on this questionnaire. It will not be possible for us to say which answers were given by which household.

The research is led by Professor Ines Meyer from the Section for Organisational Psychology. If you have any questions you can contact her on ines.meyer@uct.ac.za or 021 650 3829. It is important that you know that you can also stop participating in this study at any point.

Date of Interview_____Time Started_____Time Ended_____

Comments _____

_____

_____

_____

BLOCK 1: SOCIO-DEMOGRAPHIC PROFILE OF THE RESPONDENT

***1.* What is your age?_____** *(write down age in years)*

2. What is your gender?

Male ☐
Female ☐
Other ☐
Prefer not to say ☐

3. What best describes your ethnicity?

Asian ☐
Black African ☐
Coloured ☐
Indian ☐
White ☐
Other ☐
Prefer not to say ☐

3a. What is your home language?

Afrikaans ☐
English ☐
isiNdebele ☐
isiXhosa ☐
isiZulu ☐
Sepedi ☐
Sesotho ☐
Setswana ☐
siSwati ☐
Tshivenda ☐
Xitsonga ☐
Other ☐

4. How many family members live in your house? _____

5. How many people do you support with your salary? _____

6. What is the highest level of education that you completed? _____

7. What jobs do you have? Position(s) _____

***8.* How long have you been with your employer(s)?** *(write down the number of years)*

Employer 1 _____

Employer 2 _____

Employer 3 _____

Employer 4 _____

Employer 5 _____

9. Nature of main employment:

9a._____Permanent _____Contract _____Project-Based

9b._____Full-time _____Part-time

10. How many people are employed by your main employer?

_____ less than 10 _____10 to 49 _____50 to 149 _____150 to 499_____500 or more

11. Is there an employee union in the workplace?

_____ Yes _____No _____I don't know

12. Are you a union member?_____Yes _____No _____I don't know

13. Do you get paid **daily**, **weekly**, **fortnightly** or **monthly**? *(circle the applicable one)*

14a. If you do not mind sharing, how much income do you get paid into your bank account or in cash when you get paid?

_____

(Only ask Question 14b. if 14a not answered)

14b. Would you be willing to share in which category your monthly income **per month** falls?

Under R2 000	R2 000 to R3 500	R3 500 to R7 000	R7000 to R10 000	R10 000 to R15 000	R15 000 to R25 000	Over R25 000

15. How many days did you work in the last month? _____

BLOCK 2: PERCEIVED IMPORTANCE OF SPECIFIC DOMAINS OF LIFE

Now I am going to read a list of things in a person's life that may or may not be important to you to have a good life. For each thing that I will mention, please indicate if this is ***not at all important***, ***not really important***, ***important***, or ***very important***.

Encircle the appropriate answer:

16. How important is **HOUSING** for you to say that you have a good life? *This refers to having a place that is near your place of work, school, etc.; that it is safe from fire and floods; clean and strong, with electricity, water and toilet facilities.*	Not at all important	Not really important	Important	Very Important
17. How important is **QUALITY OF NEIGHBOURHOOD** for you to say that you have a good life? *This refers to having a safe place to live; where you have good relations with your neighbours.*	Not at all important	Not really important	Important	Very Important
18. How important is **EMPLOYMENT** for you to say that you have a good life? *This refers to having a regular, permanent and legal employment; a place to work that gives you adequate pay.*	Not at all important	Not really important	Important	Very Important
19. How important is **QUALITY OF WORKING LIFE** for you to say that you have a good life? *This refers to having a safe place to work; that is suitable to your education, and where you enjoy good relations with your boss and colleagues.*	Not at all important	Not really important	Important	Very Important
20. How important are **SAVINGS, WEALTH AND BELONGINGS** for you to say that you have a good life? *This refers to having your own house, appliances, savings and being free from debt.*	Not at all important	Not really important	Important	Very Important
21. How important are **SOCIAL RELATIONSHIPS** for you to say that you have a good life? *This refers to being with your spouse, child/children, friends, and enjoying the love of your relatives and friends.*	Not at all important	Not really important	Important	Very Important
22. How important are **LEISURE AND SPARE TIME ACTIVITIES** for you to say that you have a good life? *This refers to having time for yourself, being able to rest, relax, and have fun with your loved ones.*	Not at all important	Not really important	Important	Very Important

23. How important is **PHYSICAL HEALTH** for you to say that you have a good life? *This refers to being free from sickness and disability, being able to exercise regularly, having regular & nutritious food, enough sleep, and a long life.*	Not at all important	Not really important	Important	Very Important
24. How important are **PSYCHOLOGICAL/ MENTAL HEALTH & EMOTIONAL WELLBEING** for you to say that you have a good life? *This refers to giving importance to yourself, having a clear mind, being calm and at peace, and the ability to make personal decisions. This also refers to being respected by your family and other people; the ability to handle your problems and face changes.*	Not at all important	Not really important	Important	Very Important
25. How important are **RELIGION AND SPIRITUAL LIFE** for you to say that you have a good life? *This refers to having the opportunity to worship, pray, give to the church/synagogue/mosque/temple, and doing good to others.*	Not at all important	Not really important	Important	Very Important
26. How important are **INFORMATION AND KNOWLEDGE** for you to say that you have a good life? *This refers to having the ability to read and write, finish school, learn in different ways aside from school (e.g. watching TV or reading the newspaper), study in a good school (if you plan to study again) and being able to use your education.*	Not at all important	Not really important	Important	Very Important
27. How important is **POLITICAL PARTICIPATION** for you to say that you have a good life? *This refers to knowing what is happening in the country, voting in the elections, joining community organizations, and being free to express your political opinion.*	Not at all important	Not really important	Important	Very Important
28. How important is **GOVERNMENT PERFORMANCE** for you to say that you have a good life? *This refers to having a country that is peaceful, crime-free, has good public service and a stable economy, where citizens are united, and where there are equal justice and opportunities to everyone.*	Not at all important	Not really important	Important	Very Important

BLOCK 3: PERCEIVED FREEDOM TO ATTAIN SPECIFIC DOMAINS OF LIFE

I have asked you how important certain things are for a good life. Now I would like to ask you how possible it is for you to achieve these. For every point that I will mention, please say if for you it is **completely impossible, almost impossible, quite possible, completely possible**.

Encircle the appropriate answer:

29. How possible is it for you right now to get **HOUSING** that allows you to have a good life? *This refers to having a place that is near your place of work, school, etc.; that it is safe from fire and floods; clean and strong, with electricity, water and toilet facilities.*	Completely impossible	Almost impossible	Quite possible	Completely possible
30. How possible is it for you right now to get the **QUALITY OF NEIGHBOURHOOD** that allows you to have a good life? *This refers to having a safe place to live; where you have good relations with your neighbours.*	Completely impossible	Almost impossible	Quite possible	Completely possible
31. How possible is it for you right now to get **EMPLOYMENT** that allows you to have a good life? *This refers to having a regular, permanent and legal employment; a place to work that gives you adequate pay.*	Completely impossible	Almost impossible	Quite possible	Completely possible
32. How possible is it for you right now to get the **QUALITY OF WORKING LIFE** that allows you to have a good life? *This refers to having a safe place to work; that is suitable to your education, and where you enjoy good relations with your boss and colleagues.*	Completely impossible	Almost impossible	Quite possible	Completely possible
33. How possible is it for you right now to get the **SAVINGS, WEALTH AND ASSETS** that allow you to have a good life? *This refers to having your own house, appliances, savings and being free from debt.*	Completely impossible	Almost impossible	Quite possible	Completely possible
34. How possible is it for you right now to get the **SOCIAL RELATIONSHIPS** that allow you to have a good life? *This refers to being with your spouse, child/children, friends, and enjoying the love of your relatives and friends.*	Completely impossible	Almost impossible	Quite possible	Completely possible
35. How possible is it for you right now to get the **LEISURE AND SPARE TIME ACTIVITIES** that allow you to have a good life? *This refers to having time for yourself, being able to rest, relax, & have fun with your loved ones.*	Completely impossible	Almost impossible	Quite possible	Completely possible

36. How possible is it for you right now to get the **PHYSICAL HEALTH** that allows you to have a good life? *This refers to being free from sickness and disability, able to exercise regularly, having regular & nutritious food, enough sleep, and a long life.*	Completely impossible	Almost impossible	Quite possible	Completely possible
37. How possible is it for you right now to get the **PSYCHOLOGICAL/ MENTAL HEALTH & EMOTIONAL WELLBEING** that allow you to have a good life? *This refers to giving importance to yourself, having a clear mind, being calm and at peace, and the ability to make personal decisions. This also refers to being respected by your family and other people; the ability to handle your problems and face changes.*	Completely impossible	Almost impossible	Quite possible	Completely possible
38. How possible is it for you right now to get the **RELIGIOUS AND SPIRITUAL LIFE** that allows you to have a good life? *This refers to having the opportunity to worship, pray, give to the church/synagogue/mosque/temple, and doing good to others.*	Completely impossible	Almost impossible	Quite possible	Completely possible
39. How possible is it for you right now to get the **INFORMATION AND KNOWLEDGE** that allow you to have a good life? *This refers to having the ability to read and write, finish school, learn through different ways aside from school (e.g. watching TV or reading the newspaper), study in a good school (if you plan to study again) and being able to use your education.*	Completely impossible	Almost impossible	Quite possible	Completely possible
40. How possible is it for you right now to **PARTICIPATE POLITICALLY** in a way that allows you to have a good life? *This refers to knowing what is happening in the country, voting in the elections, joining community organizations, and being free to express your political opinion.*	Completely impossible	Almost impossible	Quite possible	Completely possible
41. How possible is it for you right now to get the **GOVERNMENT PERFORMANCE** that allows you to have a good life? *This refers to having a country that is peaceful, crime-free, has good public service and a stable economy*, where citizens are united, and where there are equal justice and opportunities to everyone.	Completely impossible	Almost impossible	Quite possible	Completely possible

BLOCK 4: ORGANIZATIONAL CITIZENSHIP BEHAVIOURS + TASK PERFORMANCE

Now I would like to ask about your experiences at work. For this part, there will be 6 choices for your answer namely

Never, Almost never, Sometimes, Often, Almost always, and **Always**

Encircle the appropriate answer:

	Never	Almost never	Some-times	Often	Almost always	Always
42. I help others who have been absent from work.	Never	Almost never	Some-times	Often	Almost always	Always
43. I help others who have heavy workloads.	Never	Almost never	Some-times	Often	Almost always	Always
44. I volunteer to assist supervisors with their work.	Never	Almost never	Some-times	Often	Almost always	Always
45. I take time to listen to co-workers' problems and worries.	Never	Almost never	Some-times	Often	Almost always	Always
46. I go out of my way to help new employees.	Never	Almost never	Some-times	Often	Almost always	Always
47. I show concern to my co-workers.	Never	Almost never	Some-times	Often	Almost always	Always
48. I share information with co-workers.	Never	Almost never	Some-times	Often	Almost always	Always
49. My attendance at work is better than that of most others.	Never	Almost never	Some-times	Often	Almost always	Always
50. I give advance notice when I can't come to work.	Never	Almost never	Some-times	Often	Almost always	Always
51. I take work breaks that are not permitted.	Never	Almost never	Some-times	Often	Almost always	Always
52. I spend much time with personal phone conversations.	Never	Almost never	Some-times	Often	Almost always	Always
53. I complain about unimportant things at work.	Never	Almost never	Some-times	Often	Almost always	Always
54. I follow unwritten rules to maintain order at work.	Never	Almost never	Some-times	Often	Almost always	Always
55. I only put minimal effort into my current job because I really feel that I'm wasting my time at work.	Never	Almost never	Some-times	Often	Almost always	Always

56. I don't know why I'm doing this work; it's pointless work.	Never	Almost never	Some-times	Often	Almost always	Always
57. I put a great deal of effort into my current job because I have fun doing my work.	Never	Almost never	Some-times	Often	Almost always	Always
58. I put a great deal of effort into my current job because what I do in my work is interesting.	Never	Almost never	Some-times	Often	Almost always	Always
59. I do little work in my current job because I don't think this work is worth putting a great deal of effort into.	Never	Almost never	Some-times	Often	Almost always	Always
60. I put a great deal of effort into my current job because the work I do is interesting.	Never	Almost never	Some-times	Often	Almost always	Always
61. Overall, I am content with my work.	Never	Almost never	Some-times	Often	Almost always	Always
62. I am content with the important aspects of my job.	Never	Almost never	Some-times	Often	Almost always	Always
63. I like the work that I do in my job.	Never	Almost never	Some-times	Often	Almost always	Always

BLOCK 5: EMPOWERMENT + FAIRNESS AT WORK

In this section of the survey, we will use a different scale to almost every question.

64. Empowerment at work means you feel you **have some control over how you do your work**, and in workplace decisions that directly affect you. On a scale from 0 to 10 how 'empowered' do you feel at work in general? 0 means not empowered at all, 10 means completely empowered

*(*Please indicate *how empowered you feel. 10 is full empowerment and 0 is zero.)*

0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐

65. I am satisfied with my job in general.	Never	Almost never	Some-times	Often	Almost always	Always
66. I am satisfied with my pay.	Never	Almost never	Some-times	Often	Almost always	Always
67. My work makes me feel proud.	Never	Almost never	Some-times	Often	Almost always	Always
68. I am stressed.	Never	Almost never	Some-times	Often	Almost always	Always
69. I am physically well.	Never	Almost never	Some-times	Often	Almost always	Always
70. I am satisfied with life in general.	Never	Almost never	Some-times	Often	Almost always	Always

71. Overall, how fair would you say your pay is?	Very Unfair	Unfair	Quite Unfair	Neither Unfair nor Fair	Quite Fair	Fair	Very Fair

Do you think your wage is a 'fair rate'…			
72. For the job?	Yes	No	Don't Know
73. Compared to your effort?	Yes	No	Don't Know
74. Compared to your qualifications?	Yes	No	Don't Know
75. Compared to similar jobs elsewhere?	Yes	No	Don't Know
76. Compared to <u>your direct supervisor</u> in the organization?	Yes	No	Don't Know

77. Does your pay provide enough for your basic needs?	Yes	No
78. Does your income exceed what you consider to be your basic needs? (e.g., it enables you to have some savings, to enjoy some leisure activities, to enjoy the occasional treat)	Yes	No

79. Do you feel that your employment income contributes not only to your own life but also to others in your household (parents, dependents)?	Yes	No	Not Applicable

BLOCK 6: LIVING ARRANGEMENTS

We now have just a few more questions about your household.

80. What best describes your living arrangements?

Rent home (formal structure) ☐
Rent home (informal structure) ☐

Home owner (formal structure, **with bond**) ☐
Hone owner (informal structure, **with bond**) ☐

Home owner (formal structure, **bond free**) ☐
Home owner (informal structure, **bond free**) ☐

81. How many children live in the household?

1 ☐ 2 ☐ 3 ☐ 4 ☐
5 ☐ 6 or more ☐

82. What is the total number of (working) incomes in your household?

1 ☐ 2 ☐ 3 ☐ 4 ☐
5 ☐ 6 or more ☐

83. How many of these incomes are full-time?

1 ☐ 2 ☐ 3 ☐ 4 ☐
5 ☐ 6 or more ☐

84. How many other dependent adults live in your household?

1 ☐ 2 ☐ 3 ☐ 4 ☐
5 ☐ 6 or more ☐

To the interviewer:
Please record the end time on the first page of the questionnaire now.
Then hand the questionnaire to the participant so they can complete the last page by themselves and place the questionnaire in an envelope when done.

BLOCK 7: INTERVIEWER CHARACTERISTICS

This part is for you to complete yourself.
Please be as honest as you can.
All responses are kept anonymous from the interviewer.

85. Overall, how would you rate your experience with your interviewer?

Very poor	Poor	Quite poor	Neutral	Quite good	Good	Very Good

86. Overall, how would you rate the **physical appearance** of your interviewer?

Very poor appearance	Poor appearance	Quite poor appearance	Neutral	Quite good appearance	Good appearance	Very good appearance

87. How would you describe your interviewer?

Please write down the first five things that come to your mind in any language you want.
There are no right or wrong answers.

(1) _____

(2) _____

(3) _____

(4) _____

(5) _____

If you would like to give any further comments about the study you can write them here:

This is the end of the questionnaire. Thank you very much.
Please place the questionnaire in the interviewer's envelope.