CONSTRUCT REPRESENTATION OF SELF-REPORT FUTURE TIME PERSPECTIVE FOR WORK AND RETIREMENT SCHOLARSHIP

A Dissertation Thesis

Presented to

The Academic Faculty

By

Matthew James Kerry

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in Psychology

Georgia Institute of Technology

May, 2015

Construct Representation of Self-Report Future Time Perspective for Work and Retirement Scholarship

Dr. Susan E. Embretson, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Jack M. Feldman
School of Psychology
Georgia Institute of Technology

Dr. Audrey Duarte
School of Psychology
Georgia Institute of Technology

Dr. Rustin Meyer
School of Psychology
Georgia Institute of Technology

Dr. David M. Sluss
College of Management
Georgia Institute of Technology

Date Approved: December 10th, 2014

For Larry

For Kathrin

For Ludo

ACKNOWLEDGEMENTS

I must first thank my advisor, Susan Embretson, for the immense encouragement, patient mentoring, and steadfast support rendered to me in a delicate time. I am grateful for the provocative engagements as much as for the congenial conversations. I am further humbled and privileged as a student to the field of psychometric assessment – innervating me to rediscover rational inquiry-*cum*-legitimate empiricism in the social sciences. A new plane for practice and research. I am grateful now and always - thank you.

I would also like to acknowledge and give special thanks to my committee members – Drs. Rustin Meyer, David Sluss, Jack Feldman, and Audrey Duarte. All of these members have influenced my maturation as a scientist and the development of this thesis.

To my family, thank you for seeing me through. To my mom, for your unwavering compassion. To pops, for instilling that I can always leave something better than I found it while being exactly where I should be. To all my brothers, Johnny, Stephen, Allie - I am proud to be a compilation of all your strengths.

To Kathrin. Thank you for staying (again). We have much ado about anything.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

$1 - \beta$.. Estimated Statistical Power
AIC.. Akaike Information Criterion
ALP………………………………………………………… American Life Panel
α... Estimated Type-I Error Rate
BIC.. Bayesian Information Criterion
β.. Sample Standardized Regression Weight
C-FTP……………………………………….... Carstensen's Future Time Perspective
C.I... Confidence Interval
Cronbach's α.. Estimated Internal Consistency Reliability
CV……………………………………………………………… Construct Validity
d.f. .... Degrees of Freedom
DML…………………………………………..……Direct Maximum Likelihood
E-H……………………………………….. Hypothesis Tested in Experimental Study
F.. Test Statistic for Analysis-of-Variance
f... Estimate of Effect Size for Analysis-of-Variance
FTP……………………………………………………….. Future Time Perspective
GLS…………………………………………………….... Generalized Least Squares
H-FTP…………………………………………… Hershey's Future Time Perspective
IRT…………………………………..………………… Item Response Theory
j……………………………………………………….………… Indicator Variable
k…………………………..…………………………………….….. Latent Variable
k_i.. Number of Item Response Categories
M... Estimated Sample Mean
MAR.. Missing-at-Random
MCAR... Missing-Completely-at-Random
McDonald's ω... Estimated Scale Homogeneity
ML………………………………………………………… Maximum Likelihood
N .... Total Sample Size
n………………………………………………………….... Individual Participant
NI .... Non-Ignorable Missing Data Pattern
p....................................... Probability Obtained Statistical Result Occurred by Chance
r .... Sample Observed-Correlation Coefficient
R-H…………………………………Hypothesis Tested in Retro-Observational Study
Retro-Obs………………………………………… Retrospective-Observational Study
RMSEA.. Root Mean Square Error of Approximation
SD .... Standard Deviation
SE... Standard Error
SEM…………………………………………………..... Structural Equation Modeling
SLE…………………………………………………….. Subjective Life Expectancy
TP…………………………………………………………………… Time Perspective

SUMMARY

The dissertation presents evidence on the measurement properties of self-report items in contemporary organizational contexts (Podsakoff & Organ, 1986). Operationally, the dissertation adopts a construct representation approach to construct validity, defined by the response processes engaged for measurement performance in trait assessment (AERA, 2014; Embretson, 1983). For example, self-report measures are known to be affected by a variety of variables, such as semantic and referent features (Cermac & Craik, 1979; Kelly, 1955) and design factors that impact cognitive context (Stone, et al, 2000; The Science of Self-Report). In turn, the response processes impacts the external correlations (Embretson, 2007). To the extent that semantic-referent features and design factors are construct-irrelevant, reduced external correlations can be expected. This dissertation presents evidence from a qualitative review of self-report future time perspective (FTP) instruments across organizational and retirement contexts. A quantitative review compares external correlates of the two instruments. A retrospective-observational study benchmarks the psychometric properties of Carstensen's self-report instrument using modern latent-variable modeling (item-response theory [IRT]). Structural equation modeling (SEM) is further used to test for moderating effects of subjective life expectancy (SLE) on latent predictors of FTP and retirement plans. Evidence from a '3 x 2' mixed-subjects experimental design is also presented indicating the effects of subjective life expectancy (SLE) on measurement error in personality factors, FTP, and retirement plans. Discussion centers on advancing measurement paradigms in psychological and education research, as well as -more generally- adopting an integrated perspective of construct validity for advancing and evaluating substantive research.

INTRODUCTION

Despite its consistency with current test standards, the centralization of construct validity (CV)[1] in test validation procedures is not uncontested. Arguments against the framework include, impractical complexity, scarce precise theories, and formal inclusion of test consequences, according to Kane (2013). However, withstanding arguments against the unifying framework, the current investigation is partly motivated by claims as to the viability of modern test theory standards for psychological practice. "[CV] has been useful as a unifying framework on a theoretical level, but has not, in itself, been an effective unifying influence on an operational level" (Kane, 2004, p. 140). If corrected-operational validities[2] satisfy *reductico ad absurdum,* this author takes chances with a unified CV framework. To this end, the current dissertation presents a practical imperative wanting a stronger program for validation research than historically conceived in functionalist paradigms (Cronbach, 1971; 1989).

Extant validation procedures, typically, subscribe to descriptive and *in vivo* approaches, that is, either new measures are developed with little theoretical scrutiny or, otherwise, existing measures are taken as indicators *prima facie*, whereby, external validity evidence inherently confers construct meaning (see, Embretson, 2007). Sometimes, constructs are re-conceptualized to conform to aberrant external correlates (c.f., Reckase & McKinley, 1991). That is, obtained external evidence

---

[1] CV will refer to both “construct validity” and “construct validation” for the remainder of this dissertation.

[2] Operational validity defined in the present context as observed-correlation coefficients corrected for criterion-instrument reliability and predictor-indirect range restriction, as protocoled in Hunter, Schmidt, & Le, 2006). Conceptually, operational validity equates measurement with meaning. See, however, an excellent commentary on bias amplification resulting from multivariate correction (Pearl, 2012), principally similar to Spector & Brannick’s (2011) exposé on misuse of statistical controls.

predisposes test meaning, rather than informing measurement accuracy.[3] Test meaning derived from external evidence, however, may be vulnerable to population differences (Haberman & Dorans, 2011), in turn, creating population-specific impact on constructs. Taken to extreme, observation-borne interpretations may lead to dismissal of test content altogether (Berg, 1955). That is, the generality of response patterns would sufficiently negate specificity of test content (c.f., Goldberg & Slovic, 1967). Conversely, rigorous empirical development of a test, e.g., item design features and test specifications, may lead to higher utility or significance in practice. Joint-standards across educational, psychological, and measurement associations advocate systematic test development for both test developers and users. The advanced-directives in test and item design raise accountability in construction and implementation, respectively (AERA, 1999; 2014).

One critical tool in construct representation research is advanced model-based measurement (see, Bennett (1990) in Frederiksen, Mislevy, & Bejar, 1993; Also, see Birnbaum in Lord and Novick (1968); Embretson, 2010). Generally, model-based measurement is premised on rationales for behavioral models in assessment. Here, model-based measurement refers to latent-trait theory as more comprehensive parameterization of observables to latent space. For example, item response theory (IRT) figures prominently in model-based measurement of modern test theory. IRT departs from classical test theory (CTT) in model specificity (item-level) and comprehensiveness of latent trait estimates, derived jointly from responses and item properties, *inter alia* (Embretson & Reise, 2000). The advanced psychometric approach has achieved state of the art in educational and psychological assessment (c.f., van der Maas, Molenaar, Maris, Kievit, & Boorsboom, 2011) . Lack of

[3] A contrapositive to Maier’s Law (1960), whereby retained factual evidence impels construct redefinition.

implementation is detrimental to advancing psychological assessment, and CV generally, by limiting applicability in real-world testing scenarios. Taken in the context of overlapping criteria between employment and retirement amid an aging population, a more precise approach to measurement, coupled with greater inquiry into its sources of error, is needed.

The remaining introduction comprises five sections. First, an overview of the construct representation approach to CV for understanding response processes in CV research is presented. In the second section, I overview the psychometric evidence justifying modern test theory for separating examinee and test - item characteristics and its faculty in construct representation research. This section will also address the comparative advantages of IRT, relative to CTT, for estimating measurement bias[4]. In the third section, I provide a compendium of the psychological research on response processes in self-reports with emphasis on testing conditions and specifications. The fourth section locates testing conditions and test specifications as two sources of validity evidence in the universal framework adopted for this dissertation. This section will also note the concomitant sources of validity evidence invoked in later chapters. The introduction concludes with a general statement of purpose and brief commentary on the inferences permitted from the design and measurement models employed in subsequent proposed studies. Chapter 2 will introduce the applied context of the focal constructs under study, future time perspective and retirement planning.

[4] Measurement bias is defined, here, as measurement scale non-invariance, resulting in differential item performance across nominally observed-groups.

1.1 Overview of Construct Representation in Validation Research

Construct representation[5] is an aspect of construct validity, whereby, multiple internal sources of validity evidence are precedential to external sources (Embretson, 1983). The approach restores principles envisaged in Cronbach and Meehl's (1955) "strong program" for validation research (Cronbach, 1989, p. 162). It also directly implicates formal cognitive theory in test development and, in service to predefined measurement goals, is subsequently estimated with model-based, modern test theory. For illustrative purposes, consider how model-based measurement may more precisely account for competing predictions, inherent to experimental designs, during incipient analytic stages to initial test development.

Conceptually, the construct representation aspect of validation may be located at the nexus of purposive theoretic measurement, domain process sampling, and assessment engineering (Leucht, 2006, also, see Bennett, 1990). Restated parenthetically, construct representation research binds "content and substantive aspects of construct validity" (Messick, 1995, *p*.745), operating at "the intersection between theory and empirical operationalization through item characteristics" (Ferster, 2013, *p*.6), and more precisely as "[CV] is explicated at the item level." (Embretson, 2008; pp.331 [abbreviation added]). As Janssen summarizes, "Construct representation is concerned with identifying the theoretical mechanisms that underlie item responses . . . The goal of construct representation is task decomposition" (2010; p. 227). Concordantly, and in contrast to CTT's tilt toward person exchangeability,

[5] The duality of "representativeness" as described by Messick (1995), then, is similar to that which Embretson (1983) articulated earlier as "reflected in" when referring to constructs associated with test scores and their interpretations. As Shepard notes, "the content-response dichotomy is merely a restatement of the logical-empirical sources of validity evidence" (1993; p.415). This author disagrees with the implied mutual exclusivity of logical and empirical 'sources' of evidence.

construct representation concerns the commonality of performance across tasks as 'relatively' more reliable estimates than between-persons.

Finally, internal sources of validity evidence, consequently, inform the nomothetic span (external evidence) aspect of validation. That is, external evidence is repurposed from defining test meaning, to indicating a test's utility as a measure of individual differences. It is also noteworthy that the strength of nomothetic span evidence may serve as indicator of the quality of test development procedures (construct representation). In this sense, the two bodies of internal and external evidence, while conceptually and empirically distinct, are informationally interrelated under the unifying framework of CV. This recapitulates the approach adopted in the conducted studies of the current dissertation.

1.2 Psychometrics of Modern Test Theory for Response Process Decomposition

With its emphasis on reliability, classical test theory maintains popularity and utility with enduring principles, but it may err in its categorical exchangeability of persons, at the expense of items or tests. Stated differently, it assumes equal measurement error across all examines, regardless of ability or location on the theoretical latent continuum, e.g., homoscedasticity. Concomitantly, by focusing on reliability at the composite-level, rather than its comprising items, it imposes parallelism to all items of the test. In other words, "items are considered to be parallel instruments" (van Alphen et al, 1994; p.197), permitting summed composites. These strong assumptions have implications for CTT's error terms, specifically, it is expressed as an additive component to the true score in estimating observed scores, as expressed by Lord and Novick (1968): $Y_{pv} = \theta_p + \varepsilon pv$.

From this formulation, a person's observed score is a summation of their true score plus a random error component. In CTT, all error is random error, that is, in the above-formulation, the error term (εpv)[6] is disassociated entirely from a person's true score (θ_p). This stringent condition, while computationally convenient, imposes serious limits to estimating error, that is, systematic or correlated error that may be related to the assessment occasion and, thereby, of likely detriment to CV. This limitation was also recognized in Cronbach's (1950) classic text on response sets, defined as temporary response biases attributable to ephemeral sources, "The writer has attempted to formulate rationally the response-set problem in factorial terms. The analysis has been unsuccessful, primarily because response sets do not obey the fundamental additive law of factor theory" (p.149), proxy, CTT.

While CTT was institutionalized as the choice measurement model for psychological research for decase, a more precise measurement system was being developed with early conceptualizations presented in Lord and Novick (1968), c.f., Suppes (1968). The advantages of the newer system begin with more granular estimation of stimuli *in situ*, that is, individual items rather than composites. In addition, contrasted to CTT, reliability is not the primary concern. This is because conditional probability estimates, coupled with varying error over trait levels, provides an information tradeoff between items and persons. That is, from inestimable error in CTT, but rather to distributed error to items, thereby enabling greater precision in measurement. Also inherent in the decomposition of error is the scalability of, both items and person parameters (see, Embretson & Poggio, 2012). Scoring models that utilize CTT measurement principles, therefore, may be scrutinized for such issues as differential validity, adverse impact, and measurement bias.

[6] See, Groves & Lyberg (2010) for a different approach to deconstructing this error from the 'sample side', also, for a general review of the history and maturation of the field of *total survey error*.

Importantly, of the aforementioned three issues, measurement bias is one that requires a precise understanding of why statistical properties should not connote meaning of substantive impact resulting from test scores. This may be illustrated by understanding that differential validity does not necessarily imply test bias. Further, tests with similar predictive validities can masquerade as fair tests, while underlying distributions of requisite knowledge, skills, or strategies for solving items differ across populations and become evident in observed mean-differences. Finally, measurement bias as indicated by differential item validity (DIF), may, in fact, be unbiased in a multidimensional, latent space (Ackerman, 1992).[7] This latter presumption, however, may require impractical precision of content domain compilation across latent levels. This will be elaborated in chapter 4.

In summary, the precision afforded by advancements in IRT-based psychometric modeling allows the decomposition of response processes. It is particularly useful in more precise modeling of would-be random error. The next section will expound on more substantive, psychological approaches to studying response processes and errors in self-report measures, followed by an introduction to sources of validity evidence, more generally, from the universal system framework adopted for this dissertation.

1.3 Psychological Research on Self-Report Response Processes

It has long been known that questionnaires, e.g., Q-data (Cattell; 1957) are vulnerable to many sources of error[8] (Cronbach, 1946; 1950; Lorge, 1937). At their

---

[7] See, Appendix A, for a more detailed, side – side comparison of CTT & IRT measurement paradigms, adopted from Embretson & Reise, 2000.

[8] More generally, Thorndike (1947) conceptually factors sources of variance in testing into general / specific and transient / lasting dimensions.

highest level of abstraction, measurement errors[9] associated with self-report data have been taxonomized into person- and method-related factors (see, Viswanathan, 2005). This distinction is also evinced in the broad (Bagozzi, 1984; Edwards, 2008) and narrow (Lance, Baranik, Lau, & Scarlau, 2009) views of general method bias, where “broad” encompasses all influential aspects of the measurement procedure, including response tendencies and context effects, while the “narrow” view excludes these aspects to refocus on ‘static’ features of the test (stimuli) itself. This dissertation technically complies with the narrower view in examining two self-report questionnaires (mono-method), but adopts a broader view for implicating the purpose of assessment in validation (Campbell & Fiske, 1959; Messick, 1989). Put simply, item and individual exchangeability are axes to be balanced in measurement inferences (Guttman, 1952; Hammond, 1948), while the current dissertation tilts toward the former.[10]

As Nolet and Tindal (1990) note, “The logic of construct validation assumes adequate domain sampling and avoids the mono-operation bias (p.20). Returning to measurement errors, sources commonly coalesce around item content-, response format-, and administration-related factors (Bagozzi, Yi, & Phillips, 1991; Bardo & Yeager, 1982; Fiske, 1982; Podsakoff, MacKenzie, & Podsakoff, 2012). In addition to these relatively static, or –internally- fixed, features of the assessment occasion, the interactive potential for construct-irrelevant sources of error, that is, between subjects and stimuli, is also well-documented (Meehl, 1945; Messick & Jackson, 1958;

[9] Sacket (1979) catalogues biases in analytic research under measurement and sampling. This may be for nominal or organizational convenience, however, the opinion of the author is this categorization emanates from the overarching paradigm of domain sampling via personological and situational approaches to the study human behavior; The postulates for both ontological and epistemological inferences are viewed as inclusive to both approaches.

[10] See, also, Zumbo (2007) or Lindley (1972) for reviews.

Jackson, 1971). Further, item parameters have been shown to vary systematically as a function of their location within a test battery or session (Leary & Dorans, 1985). Taken together, this evidence deals imperatives for the significance and interpretation of scores derived from self-reports. In short, it evinces Cronbach's "Type II" response sets, deemed "unquestionably harmful" (1950; p.18) with respect to CV.

Paulhus (2002) has defined response biases as, "Any systematic tendency to answer questionnaire items on some basis that interferes with accurate self-reports" (p. 49). As Millshap and Everson (1993) observe, "Studies of measurement bias provide empirical tests of construct interpretations" (p. 329). The mutual inclusiveness of item bias and construct validity is articulated by Ackerman (1992). While the cognitive, sequential order of response processes in self-report responses has been detailed (Tourangeau & Rasinski, 1988, see also, Holtgraves, 2004; Johnson, 1981; Stone et al., 2000), it should be noted that the taxonomization of these phenomena have been less systematic and generally diverge on account of method- and content-factor interests, as previously noted (c.f., Meyer, Dalal, & Hermida, 2010).[11] Heuristically, these within-test session phenomena, whereby "measurement changes the measure"[12] (Knowles, 1988) are termed 'context effects'. This will be elaborated in Chapter 7, but for present purposes, it is instructive to note that context

---

[11] Interestingly, in their 'Conclusion' chapter of the recent editorial volume, *New Perspectives on Faking in Personality Assessment* (2012), Maccann, Ziegler, and Roberts define faking as *intentional* response distortion. The subsequent conclusion drawn is, "The broad consensus is that faking will affect the interpretation of individual scores in high-stakes conditions but may not strongly affect interpretations of correlation-based findings (e.g., test-criterion correlations)" (*p.*313).

[12] This has origins in the observer effect in physics and, antedated, Heisenberg's *Uncertainty Principle* (1927). As it happens, recent physicists deploy a technique termed *weak measurement* to demonstrably violate Heisenberg's formulation of measurement precision-borne disturbance (Rozema et al., 2012).

effects are conjointly determined by item design principles, as well as the testing conditions themselves, including administration and instruction.

The next section will overview more specific sources of validity evidence from the universal CV framework that will be utilized for the current proposed dissertation. After this conceptual overview, I will introduce the focal concept of FTP, to which the reviewed sources of validity evidence will be invoked (Embretson, 2007).

1.4 Sources of Validity Evidence in the Universal System

In the universal system for validity schemata (see, Figure 1), sources of evidence may derive from internal and external loci, indicating construct meaning and test significance, respectively (Bechtold, 1959). Test specification is described as, “Perhaps the most essential category in determining test meaning” (Embretson, 2007, *p.* = 453). Importantly, as depicted in Figure 1, there is a feedback system from external sources of validity evidence that informs ongoing test validation efforts; the “even endless process” (Cronbach, 1989, p. 151). Also, important is the direct feedback permitted to Test Specifications as a function of external validity evidence. Hendrickson, Huff, and Luecht (2010) may aptly summarize the link between modern test theory and its relative utility for validation of test scores, subsumed by these specifications,

“conventional test specifications are usually only nominally related to psychometric properties of the items… Generating items from task models [i.e., item design features] that are ordered along the underlying performance continuum, and using those task models to control for the content, skills, and statistical properties of items, our interpretive needs (about the student) and our psychometric needs (for the test) are reconciled” (p.375). The following subsections provide brief descriptions of the sources of evidence that are invoked in subsequent study proposals.

Depicted in Figure 1, below, multiple sources of validity impact on construct meaning and test score significance (see, Embretson, 2007). The yellow and grayed figures represent the sources of validity evidence that are pertinent to the current dissertation. In particular, the two grayed sources of validity evidence will first be reviewed qualitatively in Chapter 3 under the section of internal sources of validity evidence. Typically, as protocoled by Embretson, (1998), a test developer may begin construction of a test with a specific measurement goal in mind, pursuant to some sufficiently specified theory. Chapter 3 will provide overviews of the attending theories with contradictory claims for age-related changes of FTP. I will also provide a brief review of the domain structure, and how its rigidity may be evinced in misapplication and misinterpretation of test scores.

In addition to qualitative review of these two sources of evidence (Chapter 3), I will also empirically investigate the psychometric properties of the theoretically elaborated FTP scale using latent variable analyses (IRT). I will also assess the potential for moderation of a proposed structural equation model (SEM) as a function of subjective life expectancy (SLE), which will inform subsequent proposed experiments. Differential item functioning (DIF) of FTP will also be assessed as a function of SLE, as well as retirement status. Later, the proposed experimental program will extend the focus on differential predictions of FTP to also include a rescaling procedure for an FFM measure of personality to correct for self-report biases. First, a brief statement is warranted for the proposed designs and measurement approaches of the subsequently proposed studies.

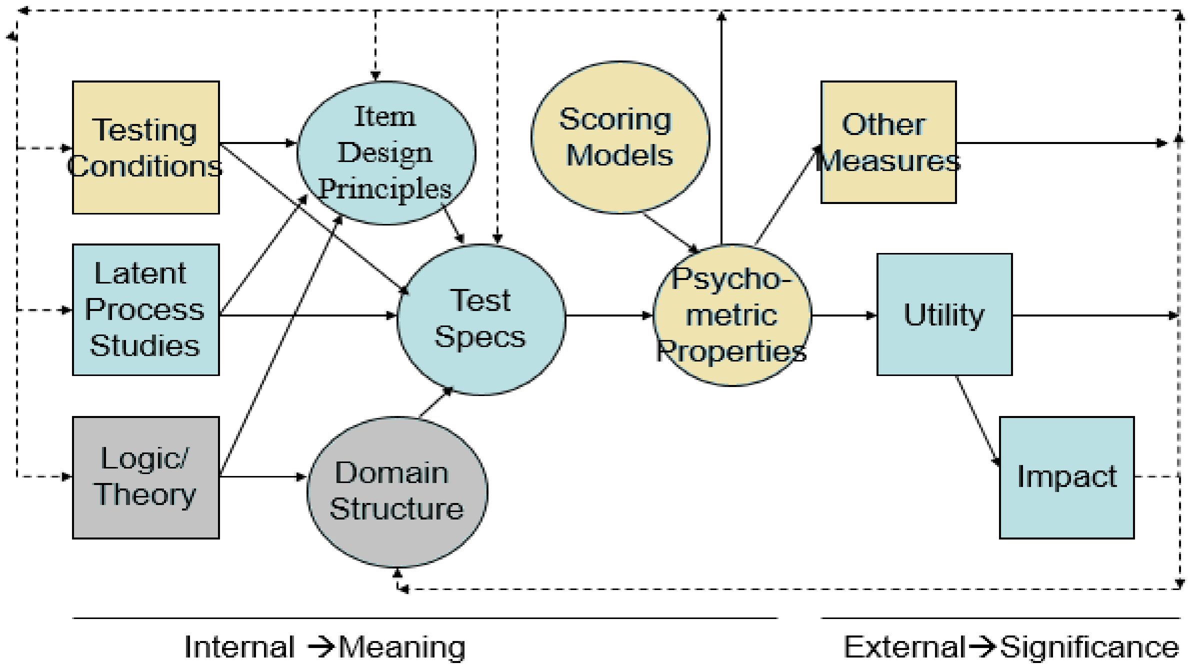


Figure 1. Sources of validity in universal system. Adopted from Embretson (2007)

1.5 Design, Measuremennt, and General Statement of Purpose

While the concept of representative design is distinguishable from construct representation research, its logic is applicable to the subjunctive of the current study. That is, it extends the notion of consistency between items, to occasions and raters as a basis for generalizability theory (Anastasi & Urbina, 1997; Cattell, 1964). The contrapositive of failing to generalize, therefore, implies an inconsistency across one or more of these three elements. Relevant for the current dissertation, the comparison of self-reports in a cross-sectional design effectively excludes occasions as a candidate source of inconsistency. To this end, the value of construct representation research, while appropriated for test development, may also function as a –gross-diagnostic indicator for sources of measurement error, that is, how much error may be attributable to item features stemming from theoretic misconception (developers → items), relative to practical misapplication (users → persons). The case of external conflicting correlates with FTP across employee and retiree populations suggest misapplication, but to no more degree than retaining two, separately constructed self-report measures[13] . In effect, the current dissertation implicates representative design to the extent that it permits convergence of evidence for determining the relatively greater contributor to measurement error.

To this end, the retrospective observational study complements the proposed experiments by isolating the effects of a single FTP measure in a heterogeneous population (consistency across items), before administering both FTP measures, concurrently, in a pilot experiment within a homogeneous population (consistency across raters). The difference in error as a function of either the population or of the measure deployed should be less informative than theoretically postulated departures from expected effects. Along with introducing the problem statement, the next chapter elaborates theorized disordinal effects (i.e., plausible rival hypotheses), which facilitates discernment of these error-driven departures. Moreover, identification of a theory-consistent condition that permits the conflicting correlates, plausibly, may serve as an explanatory variable if effectively manipulated within an experimental paradigm. For this purpose, subjective life expectancy (SLE) is identified from the literature and proposed as a potential explanatory cause of, not only directional changes in levels of the focal variable (FTP), but in its pattern of nomological correlates to relevant criteria.

---

[13] This may cavil with regard to Cronbach & Meehl's (1955) observation, "On the other hand, the accumulated evidence for a test's construct validity may be so strong that an instance of misprediction will force us to modify the subtheory employing the construct rather than denying the claim that the test measures the construct" (p. 295). This assertion is reified in Meehl's (1990) expose on appraising and amending theory, as the line between *T* (theory of interest) and A_t (the conjunction of auxiliary theories needed to make the derivation to observations go through) is blurry, i.e., hardly inverse.

Put simply, as surveyors or experimentalists, we must either randomly sample stimuli from the environment (items, occasions, or raters) or create stimuli in which identified environmental properties are preserved, c.f., naturalistic decision making, (see, for review, Dhami, Hertwig, & Hoffrage, 2004). This paradox lies at the putative antipodes of content- and criterion-related validity evidence, but the unifying framework under construct validity restores a precautionary balance. To this end, construct representation is the ontologically preventive approach, while nomological validity evidence may, in complement, promote identification of theoretic imprecision or criteria deficiency. For example, construct representation localizes the burden of test construction, indeed, the merit of criterion-related evidence to developers. Conversely, others have called for greater rigor of criteria assessment, analogously, content validity of user data (Lebreton, Scherer, & James, 2014).

To recap, methodologically, the formative interplay of construct meaning and test significance is reflected in the analytic design of the current studies. Stated differently, a construct representation approach may serve, both test development, as well test evaluation through design-guided inferences. For the current studies, the design - measurement intersection (e.g., Schwarz, 1999) will permit convergent evidence toward a specific determination. That is, a retrospective observational study employing IRT analyses will render external validity evidence, while a prospective, evaluative experiment uses internal sources of validity evidence to pit opposing predictions in the nomological network of a common measure. This contrast in design that accommodates different sources of evidence may be most succinctly summarized in Embretson's (1983) original text on construct representation, which holds currency at present:

"In experimental psychology, the main emphasis was antecedent/consequent relationships, with little interest in the intervening mechanisms that are relevant to construct representation. In correlational psychology, researchers in the factor analytic paradigm were interested in task decomposition, but construct representation was completely confounded with nomothetic span.That is, the "components" that are decomposed by principal factor analysis are based on correlations of individual differences between tasks. Unfortunately, as has been pointed out by many writers, correlations between tasks reflect many influences… The factors are "functional unities," (e.g., Thurstone, 1935) because they represent influences that cannot be separated in a given set of variables, but this does not necessarily imply that they represent elementary theoretical mechanisms." (p.180).

Indeed, neither work nor retirement are ultimate endpoints, temporally. Substantively, I argue that the relative disintegration of work and retirement scholarship is analogous to disintegrated sources of validity evidence. Future time perspective (FTP) is studied as a vessel for demonstrating the practical utility of construct representation research in validation efforts. The inauguration of a dedicated, Oxford journal at the nexus of aging-integration, organizational behavior, and retirement institutions may adjunctively beckon the state of affairs (*Journal of Work, Aging, & Retirement, forthcoming,* 2015). Overarching, then, my two goals are, 1) To provide merit for construct representation research in construct validation research, particularly in applied psychological research programs, and 2) To examine measurement properties-as-explanation for opposing predictions. Chapter 2 briefly introduces the focal construct of study, contextualized in a problem statement with attending theoretical postulates, e.g., age-related changes of FTP. Chapter 3 will more succinctly compare the theoretical frameworks and respective FTP constructs.

CHAPTER 2

WORK, RETIREMENT, AND FUTURE TIME PERSPECTIVE

Industrial / organizational (I/O) psychology's research on retirement, to date, has focused on the changing nature of retirement and its implications for older workers and retirees (see reviews by, Ekerdt, 2010; Feldman & Beehr, 2011; Shultz & Wang, 2011; Wang & Shultz, 2010; Wang & Shi, 2014). Organizational scholars have also recognized the link between the changing nature of retirement and that of work (Shultz & Olson, 2013). For example, the increasingly complex and dynamic conditions of retirement may be reflected in employers' increasing emphasis on work flexibility; the common feature is work-life balance, and the common function is retention / attraction. Importantly, the phenomenological similarities between work and retirement are more than functionally parallel, rather, I will argue that they are structurally complementary, as well. Recasting work flexibility, for illustrative purpose, one might consider how the holistic continuity of phased-workforce withdrawal complements the concurrency of work-recovery cycles (Ashford, Harrison, & Sluss, 2014; Zijlstra & Sonnetag, 2006).[14]

Casio's (1995) seminal publication on the changing nature of work helped survey some of the challenges of evolving industry, however, little attention was paid to the aging labor force. Moreover, there has been little substantive integration between psychological research on work and retirement (c.f. Hanisch & Hulin, 1990, see, Kohli, Rein, Guillemard, & van Gunsteren for review). Retirement scholars, meanwhile, have long-advocated that comprehension of the retirement process

[14] A recent, novel exposition on this relation was put forth by Sargent, Lee, Martin, & Zikic (2013), where they argue a "reinvention" of retirement could manifest as greater off-times over the conventional employment period to extend career pursuit in older age.

necessitates an understanding of employees' relative valuation of work and non-work domains (Friedmann & Havinghurst, 1954). In addition, as Wang and Shultz (2010) note, "very few studies that examined outcomes of retirement have incorporated factors that influenced the original retirement decision…This creates a logic gap" (p.176).

One recent example of a new application in adult assessment pertains to the increasing overlap of work and retirement contexts.[15] Specifically, the continually aging and mobile workforce compels precise application of lifespan motivation theories to work and retirement domains (Abraham & Hansson, 1995; Löckenhoff, 2012). Little integration has occurred, however, leaving internal sources of validity evidence of a common, bespoke measure in question (Loevinger, 1957). Unsurprisingly, substantive disconnect has led to contradictory predictions regarding the nomothetic span of a focal construct, namely, the age-related changes of future time perspective (FTP). In organizational scholarship, consistent with socio-emotional selectivity theory (SST), FTP is hypothesized to *decrease* with age (Carstensen, 1993; 1995). In opposition, a psycho-motivational model from the retirement literature predicts that FTP *increases* with age (Hershey, 2004; Hershey, Jacobs-Lawson, McArdle, & Humagami, 2007). The current dissertation aims to supply CV evidence by investigating the construct representation of self-report FTP in work and retirement assessment domains and, further, by conducting an experimental evaluation of the competing predictions. A compendium of empirical relations from the literature is

[15] This may be implied by age – period – cohort effect entanglement, that is, aging effects may be attributable to, either the persons or institutions they occupy (timescales). For example, aging effects of industry is work redesign to accommodate a cohort effect of retiring Baby Boomers, but also transmuting due to the periodic effect of the information age. In contrast, a pivot on this age effect to persons, that is longer life expectancies, is the thrust for the current dissertation, because *some* of the 'temporally' relevant accommodations for Baby Boomers will abide technologic shifts (Schaie, 1965, Masche & van Dulmen, 2004).

depicted in Table 1, below.

Table 1

Summary of Directional Empirical Findings.

Antecedents		**Mechanisms**		**Consequences**
Age	**(+)**	Retirement Planning	**(-)**	Intended Retirement
Age	**(+ / -)**	FTP	**(+)**	Intended Retirement
SLE	**(+)**	Retirement Planning	**(-)**	Intended Retirement
SLE	**(?)**	FTP	**(+)**	Intended Retirement
		Direct Effects		
Age		+		Intended Retirement
SLE		+		Intended Retirement

In terms of the self-report process, for retirement research, retirement may qualify as a ‘sensitive’ topic (Lee, 1993). For example, Krumpal’s (2013) literature review of determinants of SDR in sensitive items, Krumpal (2013) observes that non-responses is a molar indicator of content sensitivity. Extrapolating to findings from the first wave of the US HRS data, Ekerdt (1996) reports that the largest faction of respondents indicated having ‘no plans’ for retirement (~ 40%). More importantly, the incidence was constant for over 10 years (Abraham & Houseman, 2004), suggesting robustness to age and period effects and prompting Ekerdt (2001) to advocate, “Uncertainty is an authentic, meaningful stance toward retirement” (*p.* 168).

Theoretically, from organizational behavior and career development domains, Granrose and Baccili (2006) contend that the boundaryless and protean career models are indistinguishable and more generally reflect the change in employer – employee relationships. In addition, early explanatory models of retirement were, previously, subsumed as a form of voluntary turnover (Hanisch & Hulin, 1991, c.f., Horn, Mitchell, Lee, & Griffeth, 2012, Lee, Burch, & Mitchell, 2014). If the form or structure of, both contemporary work and retirement has become more complex, then there may be commensurately complex antecedents for both, as well, e.g., Hanisch (1995). However, Ekerdt (2004) aptly observed, “Retirement may become de-standardized as to the incidents of timing and form, but not as to its eventuality” (p. 6). To conclude this chapter, I will preview the dissertation’s goals appearing in subsequent chapters, followed by a statement of purpose.

First, I will introduce the two specific conceptualizations of FTP. Second, I will review two internal sources of validity evidence, namely, a theoretical analysis and examine its domain structure. Third, I will propose a retrospective, observational study of the psychometric properties of Carstensen’s (1996) FTP scale (the first conducted to this author’s knowledge). Fourth, I will overview external sources of validity evidence (nomothetic span) with emphasis on three variables that are pertinent to the subsequently proposed experiments (chronological age, retirement planning, and intended retirement). The review will also include personality correlates of FTP for purpose of the study’s auxiliary measurement goals, that is, rescaling with anchoring vignettes to correct for self-report bias. Fifth, I will introduce a prospective, between-subjects pilot experiment with survey methodology serving two primary objectives: 1) To investigate the measurement properties of the combined FTP items, the proposed retirement planning scale, and the anchoring vignettes for rescaling self-report data, and 2) To investigate the effects of test condition manipulation for evaluating changes in the level of the focal construct. Last, I will introduce the primary, between-subjects experiment in order to further evaluate the competing predictions for the developmental function of future time perspective with a larger sample.

2.1 Specific Statement of Purpose

I will use an experimental manipulation expected to differentially impact future time perspective to demonstrate changes, both in the level of the target construct under measure but, in extension, the external correlates in its nomological net. This is based on research findings that indicate item design features may predict, both the location and the discrimination parameters of test items (Embretson, 1999).

In complement, personality has also been known to be influenced by within-administration 'contextual' factors (Millar & Tessar, 1986; Tourangeau & Rasinski, 1988)[16]. Specifically, it is known that individuals' perceptions of adjectives and behaviors can be highly variable across administrations, instructions, item wordings, and response formats (see, Schwarz, 1999). One way to increase objectivity is to train participants on the stimuli being assessed. Another approach that has been offered recently is the use of anchoring vignettes for correcting participant responses, specifically, their mapping of latent responses to the physical apparatus, e.g. self-report formats. I will evaluate this latter approach of correcting self-report personality for its improvement in external validity (correlates) with FTP, as well.

These goal will be accomplished primarily through three objectives: First, I will review FTP's theoretical conceptualization and extant empirical evidence. Second, I will conduct a retrospective, observational study investigating the measurement properties of Carsten's FTP scale and latent structural relations with external variables. Third, I will conduct two between-subjects experiments testing the competing hypotheses regarding age-related changes of FTP.

---

[16] Bogacz, Brown, Moehlis, Holmes, & Cohen (2006) present a typology of sequential sampling of information models as a subordinate to superordinate item response process models.

In addition to formal hypothesis testing, three sets of exploratory procedures / analyses will be conducted, specifically, 1) The rescaling of self-report personality with anchoring vignettes to increase validity of self-report correlates, 2) Comparative effect size estimates to advance precision of lifespan motivation theories (Platt, 1964), and 3) Computation of joint-variances to provide explanation for proposed and observed effects (Schoen, DeSimone, and James,2011). The next chapter introduces the focal construct of this dissertation, future time perspective.

CHAPTER 3

FUTURE TIME PERSPECTIVE

The formal introduction of *time perspective* (TP) into psychological literature can be sourced to Lewin's (1951)[17] life space model where time perspective is defined as, "the totality of the individual's views of his psychological future and psychological past existing at a given time" (p. 75). Futural definitions of time perspective (F-TP) include, "timing and ordering of personalized future events" (Wallace, 1956), the "general concern for the future" (Kastenbaum, 1961), and the "general capacity to anticipate, shed light on, and structure the future" (Trommsdorff, 1983). Comprehensive coverage of the myriad approaches to studying FTP is beyond the scope of the current dissertation[18], however, a synopsis of its early scholastic development is useful for introducing the applied paradigms in work and retirement scholarship.

3.1 Early Scholastic Development

It is important to note that early development of FTP progressed, mostly, independent of its predecessor, time *perception[19]* (see for review, Wallace & Rabin, 1960), and this initial divide has rarely been abridged (cf. Feifel, 1957; Lomranz, Shmotkin, & Katznelson, 1983). The term *time perception*, as used here, is retained

[17] Heidegger (1927) and Frank (1939) may be credited with earlier expositions on time perspectives for
understanding human behavior and, even earlier discussion is found in a dedicated chapter, "Time Perception", by William James (1890). Also, psychophysical studies of time perception and estimation may be dated to the earliest studies of individual differences in reaction time.

[18] For an excellent discourse on the interdisciplinary study of *chronosophy* (knowledge-of-time), see Fraser, 1967. Also, for edited volumes, see Roeckelein (2000) and Grondin (2008).

[19] For taxonomical treatment of time perception, see Pöppel (1989). For a molar, graphical schematic of time phenomena, see Locsin (1993).

from the historical context from which it is derived, that is, time perception encompassed, both *time awareness* (formative assessment of subjective time) and *objective time* (calculation-estimation performance). These two “modes” would later be formally differentiated by Fraser (1966, 1975, 1978), in addition to a third “mode”, *time perspective* (reflective assessment of subjective time). Due to the ‘refracted development’ of psychological study on temporality, *time perception* is also misappropriated for studies of *time perspective*. As will be revealed shortly, the misappropriation in application goes beyond labels and, sometimes, leads to opposing predictions. Still, importantly, both qualitative and quantitative approaches applied in the early study of *time perception* and *time perspective*, respectively, converged on evidence suggestive of a developmental phenomenon related to the lifespan.

3.2 Overview of Socioemotional Selectivity Theory

A corpus of research that set-out to explore psychological mechanisms for observed-social contact reduction across the lifespan would lead to the discovery of multiple, unique developmental phenomena (Carstensen & Frederickson, 1992; Frederickson & Carstenson, 1990). These pattern of findings for the relative salience of different goals (content) over the lifespan serve the foundations of socioemotional selectivity theory (SST; Carstensen, 1993; 1995). Two primary goals, information-seeking and emotion-regulation, have been identified as particularly influential in adults’ social motives and preferences. While chronological age was originally posited as a reasonable proxy for, both past experience and future construal, Carstensen (1995) notes, “other externally imposed constraints on time, will also influence the salience of different goals” (p. 153). As a result of a limited perceived-future, it is argued that persons will prefer affectively rich interaction for the purpose of emotion regulation. On the other hand, open perceived-futures should result in information- and novelty-seeking social behaviors. An oft-neglected goal function that is located between emotion-regulation and information-seeking, in older age, is self-concept. Carstensen and Lang (1996) would later develop a 10-item questionnaire to measure future time perspective, and this is the instrument that will be included in the subsequently proposed experiments.

3.3 Overview of Psychomotivational Model of Retirement Planning

Hershey and colleagues (2000; 2004; 2007) constructed a conceptual model of financial planning for retirement, based loosely on Friedman and Scholnik's (1997) global, process-model of planning. According to the model, investment behavior for retirement is determined by four *qualitatively* different "sets": 1) psychological, 2) task characteristics, 3) cultural ethos, and 4) financial resources (Hershey et al., 2007). The psychological factor was delineated by personality, motivational, and cognitive factors. FTP, as conceptualized by Hershey and colleagues is located within the personality factor. In addition, it should be noted that Hershey adopted a hierarchical paradigm for motivation and personality (Mowen, 2000). According to Hershey (2000), *H-FTP* was also located at the "central" level of personality, between cardinal and surface traits. The construction of the scale followed from a literature review on FTP, where it was concluded, "none of the published scales adequately captured the construct" (p. 690). The exact number of items and content of the scale is difficult to ascertain, however, due to changes in the scale items across studies. A side-side comparison of the two FTP scales, organized by construct conceptualization, attending theoretical postulates, and item content is provided in Table 2.

3.4 Internal Sources of Validity Evidence

While many recent efforts have been made to recast construct validity in psychological assessment (Bornsboom, Mellenbergh, & Heerden, 2004; Cizek, 2012;

c.f. Embretson, 2007; Lissitz & Samuelson, 2007), the current dissertation adopts the basic terminology as used by Messick (1989; 1995) to describe two threats to construct validity. Specifically, *construct underrepresentation* refers to a narrow assessment that fails to include important dimensions or facets that is pertinent to the focal construct. *Construct-irrelevant variance* refers to a broad assessment that introduces unnecessary, systematic (reliable) variance that is impertinent to the focal construct. Conceptually, the two threats loosely relate to criterion deficiency and contamination, respectively.

3.5 Theoretical Analysis

Both the *C-FTP* scale from socioemotional selectivity theory and the *H-FTP* scale from the psychomotivational model of retirement planning are sourced to lifespan paradigms. Predating the maturation of lifespan theories, however, developmental theories were galvanized by their focus on different phases of life, and this difference may permeate the theorizing of future time perspective by Carstensen and Hershey.

Socioemotional selectivity theory emanated from gerontological scholarship, whereas the process theory of planning was conceptualized within early childhood development paradigms (cf. Friedman & Scholnick, 1997). Empirically, research findings on the lifespan trajectory of a variety of mixed-FTP measures indicate an inverted-U function over the lifespan. Figure 1 depicts how the relative truncation of one of these extreme segments of the population could contrive a positive or negative linear relationship.[20] For example, if each researcher were interested in studying age-

[20] In a recent review of response styles in survey research, VanVareanbergh & Thomas (2013) summarize similar contrasting findings between age and extreme-response styles (increases and decreases), before revealing that the misprediction was sourced to sampling error, specifically, that both younger and older respondents

related population effects (excluding period and cohort effects), a gerontological retirement scholar interested in, say, adaptation and SWB may not be so concerned about initial workforce entry. Conversely, an organizational behavior scholar interested in, say, affective commitment and CWB, may not be so concerned about bridge employment. This example is oversimplified and error-prone, but it is also instructive for illustrating the integrity of conflicting predictions. This example also demonstrates that random sampling methods are subject to bias when sampling frames are mispecified (how study sample is identified from the target population) (Heckman; 1979).

As an aside, it is also interesting to note that the function of future time perspective seems to –roughly- reflect the inverse of the classic-U function found for delay-of-gratification tasks over the lifespan.**[21]** All item content may be found in the bottom section of Table 2, below.

3.5.1 Hershey's FTP Scale (H-FTP). In the original, 4-item H-FTP scale, there is evidence for construct-irrelevant variance with regard to active, long-term goal pursuit (item-2). In addition, item-4 appears to relate to abstraction – concreteness perception. The mere scarcity of remaining items in the older scale may also constitute construct underrepresentation, although this is only conjecture. Procedurally, it should be noted that the original scale comprised nine items and was reduced after iterative exploratory factor analyses until a single-factor solution was obtained (see, Hershey & Mowen, 2000).

---

exhibited higher extreme-response styles, relative to middle-age adults (De Johg, Steenkamp, Fox, & Baumgartner, 2008).

[21] In their literature review of the relationship between time perspective and delay discounting, Teuscher & Mitchell (2011) report direct-comparison correlations < r = .20, suggesting separable phenomena.

In the newer, 6-item H-FTP scale, two items (1 & 4) are transported, with the addition of four new items (the remaining two items from the original scale are omitted without explanation). In this newer scale, there is also evidence of construct-irrelevant variance with regards to appeal to discretionary income practices (item-1), as well as general rationalization for planning (item-3). Items two and six seem more appropriate to the theoretical conception of FTP, specifically, as evaluative judgments (predilections) for thinking about the future and present, respectively. Finally, item five may be described as more of a behavioral indicator (proclivity) for spending time.

3.5.2 Carstensen's FTP Scale (C-FTP)[22]. In the C-FTP scale, item-3 may be described as an extremely worded item, which has been shown to cause DIF at the response option-level, as well as the item- and test-levels of analyses (Nye, Newman, & Joseph (2011). In addition, the three-reverse-scored items of the C-FTP scale are bundled at the end of the scale. Unless item-randomization is taken into researchers' discretion of scale administration, there is empirical evidence for item-order effects within a single assessment occasion that impacts item parameters (Hayes, 2012; Knowles, 1988; Millar & Tessar, 1986). In addition, item-8 conflates evaluative *prospection* with the subjective perception of real-time (e.g., *speed*).[23]

[22] Carstensen and Hershey's scales of future time perspective will be denoted as C-FTP and H-FTP, respectively.

[23] It should be noted that, neither Hershey nor Carstensen provide sufficient detail to the *construction* of their respective FTP measures, least empirically, to make clear determination of sampling domains. A contrast may be made, however, in that Hershey provides *some* empirical evidence of test construction, though there are disconcerting cases of commission and omission of evidence within, and across studies, respectively. In contrast, the original, unpublished manuscript from Carstensen and Friedman (1996) on FTP has been confirmed by its first author to have been, "developed on purely rational and theoretical bases". In consonance with the current dissertation, it should also be noted that the referent author indicated explicit support for, not only, "disconfirmatory information, not findings that support

3.6 Domain Structure

Domain structure is expected to change over time, concomitantly with theory itself (Embretson, 2007). Evidence for changes may derive from subject matter expert consensus or, alternatively, from a sufficiently detailed nomological network, wherein, refutation may be adequately accessible (Cronbach & Meehl, 1955; Meehl, 1990). Indeed, this is the very "strong program" of research advocated by Cronbach (1989). If, however, domain structure remains stagnant amid changing –operational-criterion, then score interpretations from assessments built on antedated theory may become suspect. As Kane (2001) comments, "it could be argued that criterion-based validation works best if the criteria are accepted at face value" (p.9). Extending this logic, one could argue that the "strong program" of CV research necessitates strong theory, but that weak theory may be borne by overlapping, if contradictory, nomological networks (Cronbach, 1988, p.12; Meehl, 1990).

Regarding the focal questionnaires, in contrast to the H-FTP scale, recent measurement development of C-FTP has provided evidence for its context-sensitivity (Coan, 1964; Guilford, 1961). For example, Zacher and Freese (2009) has investigated the empirical differentiation of general and occupational FTP. In addition, Zacher and colleagues (2009; 2010) report the influence of job design on occupational FTP, interpreting that favorable job conditions may elongate one's occupational FTP. Because of investigators' reliance on observational study designs, however, the precise mechanism remains inconclusive. For example, it may also be argued that favorable job conditions for elder workers does not increase occupational time horizons but, instead, may operate on staying in the work force by decreasing waiting costs.

---

hypotheses" with regards to FTP, but socio-emotional selectivity theory (SST), generally. (personal communication, July 19, 2013).

3.7 Psychometric Properties

An inspection of Table 2 reveals that differences in scale construction, item-response format, and item content all point toward the potential for measurement bias. In addition, primary studies of, either FTP scale, have rarely adopted a longitudinal or experimental study design, although relatively stronger (deductive) designs have been implemented for C-FTP, relative to H-FTP. The psychometric properties of both scales regardless of study design, however, have been limited to classical test theory frameworks, thus, the current dissertation proposes an IRT, model-based assessment of the C-FTP scale.

It should also be noted that the FTP instrument similarly format a 7-point likert-type scale with inexhaustive descriptive anchors at minimum and maximum values. Also, only three studies, to this authors' knowledge, as explored the dimensionality of the FTP instrument, and all efforts were conducted in exploratory paradigms.

Table 2

Comparison of Two Future Time Perspective Questionnaires: Construct Conceptualization, Theoretical Postulates, and Item-Content.

H-FTP	**C-FTP**
Construct Conceptualization and Definitions	
Conceptualized as 'central' trait within personality hierarchy, between cardinal and surface traits (Buss, 1989). Defined as, "patience or planning horizon…disproportional focus on future events" (Hershey, 2004, p. 34). "Designed to tap the extent to which individuals enjoy thinking about and planning for the future" (Hershey, 2007, p. 30).	Conceptualized as malleable, cognitive-motivational construct. Defined as, "perceptions of the future as being limited or open-ended" (Lang & Carstensen, 2002, p.125). Defined as, "Perception of one's future time as expansive and full of opportunities versus limited with few remaining opportunities" (Cate & John, 2007).
Theoretical Postulates	
FTP has *positive* relation with age.	FTP has *negative* relation with age.
Item-Content	
(Hershey & Mowen 2000) Future Time Orientation 1. I enjoy thinking about how I will live 10+ years in the future. 2. I have established long-term goals and am working to fulfill them. 3. It is very hard for me to visualize the kind of person I will be 10 years from now. (r) 4. The future seems very vague and uncertain to me. (r) (Hershey et al. 2007) Future Time Perspective 1. I follow the advice to save for a rainy day. 2. I enjoy thinking how I will live years from now in the future. 3. The distant future is too uncertain to plan for. (r) 4. The future seems very vague and uncertain to me. (r) 5. I pretty much live on a day-to-day basis. (r) 6. I enjoy living for the moment and not knowing what tomorrow will bring. (r)	(Carstensen & Lang, 1996; Carsten 2006) Future Time Perspective 1) Many opportunities await me in the future 2) Most of my life still lies ahead of me 3) My future seems infinite to me 4) I expect that I will set many new goals in the future 5) My future is filled with possibilities 6) I could do anything I want in the future 7) There is plenty of time left in my life to make new plans 8) I have the sense that time is running out (r) 9) As I get older, I begin to experience time is limited (r) 10) There are only limited possibilities in my future (r)

Note. Items followed by (r) indicate reverse-scored items. *H-FTP* = Hershey future time perspective, *C-FTP* = Carstensen future time perspective.

CHAPTER 4

EXTERNAL SOURCES OF VALIDITY EVIDENCE

Scriven (1987) has labeled the "use of a correlate . . . as if it were an explanation of, or a substitute for, or a valid evaluative criterion of, another variable" as the *fallacy of statistical surrogation* (p. 11). The fallacy involves a "substitution of a statistical notion for a concept of a more sophisticated kind such as causation or identity" (Scriven, 1987, p. 11). Because of the explicit comparison of two scales that purport to measure the same latent construct, the nomothetic span evidence presented below (see, Table 3) may be likened to Campbell's (1960) notion of *nomothetic validity*. Still, the empirical correlates within the scales' nomological network provides evidence on the theoretical specificity or test development quality, or both (Embretson, 1983). The magnitude of the following effect size estimates are interpreted heuristically according to Cohen's (1988) rubric[24].

4.1 Age Correlates

As shown in Table 3, the age-related correlates of the *C-FTP* scale are consistently in the theoretically postulated direction, with magnitudes ranging from fairly small (r = -.18) to rather large (r = -.70). In addition, the C-FTP scale has generally been subjected to more empirical investigation (replication) than the *H-FTP* measure. On the other hand, the age-related correlates of age with *H-FTP* are few in number and negligible in magnitude. Furthermore, some correlates run counter to the direction postulated by Hershey and colleagues (2000; 2007). It should also be noted that for the *H-FTP* scale, in particular, many published studies do not report correlation statistics. In addition, statistical evidence for positive age-related

[24] According to Cohen (1988), observed correlations of Pearson's r = .2, .5, and .8 indicate small, medium, and large effect estimates, respectively.

correlates with *H-FTP* derive from, either ordinary least squares (OLS) or structural equation modeling (SEM) analyses. Departures from the correlation coefficients as a result of these analyses may be caused by many factors, for example, multicollinearity and non-scalar invariance, i.e., scale artifact (see, Embretson & Poggio, 2012).[25]

4.2 Retirement Planning Correlates

Perhaps unsurprising, there is more evidence of retirement planning's relation with *H-FTP*, relative to *C-FTP*. The overall number of primary studies, however, is negligibly low (K = 4 : 1, respectively). For *H-FTP*, all correlates are in the theoretically postulated direction and range from rather small (r = .10) to medium in magnitude (r = .50). Interestingly, for the *C-FTP* scale, the single effect size estimate is also in the same direction with similar small-medium magnitude (r = .32). This is also interesting in context of the middle-age adult sample (M_{age} = 47.2). Fisher's r-z transformation was conducted in order to compare population effect size estimates across the two FTP scales (r-$_{\text{CFTP}}$[26] = .38 and r-$_{\text{HFTP}}$ = .33). Results indicated non-significance ($z_{(1)}$) = 1.16, p = .246), suggesting that the mean-observed effect size for *H-FTP* and retirement planning is comparable to that for *C-FTP* and retirement planning, assuming a homogeneous population. Interestingly, retirement planning has been shown to increase exponentially with proximity-to-retirement. Contrarily, intended retirement is also shown to be delayed exponentially with chronological age. This raises an important question as to, whether retirement planning is merely a precursor to, or actual accelerator of, intended retirement.

---

[25] Of particular peculiarity, however, is the referencing of Carstensen and colleagues' work in studies hypothesizing positive effects age effects on FTP (see, Hershey, Henkens, & Dalen, 2010).

[26] Fisher's r - z transformed average of observed study correlations.

4.3 Intended Retirement Correlates

For correlates with intended retirement date, there is only one study found for *H-FTP*, however, there are no postulates theorized for *H-FTP* and intended retirement date. Still, a small negative effect estimate is reported (r = -.25). For *C-FTP*, however, small positive effects are observed, ranging from r = .09 - .32. A negligible, negative correlation with intended retirement date is also reported for the *C-FTP* scale (r = -.04).

There are a few general noteworthy points with regards to the external sources of validity evidence of the two FTP scales. First, while neither paradigm explicitly incorporates workforce exit, SST theory details that "anticipated social endings" is the precipitator of future time perspective. In addition, *H-FTP* has been exclusively assessed with self-report questionnaire. In SST, however, various methods for assessing *C-FTP* predated the construction of the self-report inventory (e.g., card-sorts, sentence-completions, line-drawings). Still, there has been almost no convergent validity assessed across methods in a given study (monotrait-heteromethod). In one recent exception, DeMeyer and Raedt (2013) locally developed a scrambled sentence assessment of FTP in a college sample, and it correlated r = .48 with the *C-FTP* self-report. A follow-up study with a similar study using an imagery simulation to manipulate FTP demonstrated a correlation of r = .58 between the scrambled sentence task and *C-FTP* self-report. Chapter 6 introduces the experimental program of research for the current proposed dissertation.

Table 3

Comparison of External Correlates of Two Future Time Perspective Questionnaires.

H-FTP				**C-FTP**			
Age							
Study	**Sample (*N*)**	**Design**	**ES *r***	**Study**	**Sample (*N*)**	**Design**	**ES *r***
Hershey et al. (2007)	Full-time employees (*265*) Age$_M$ = 36.3 (6.18) [25-45]	X	-.03	Schwall (2008)	Community adults (*N* = 233) Age$_M$ = 47 (10.6)	X	-.37**
Hershey & Mowen (2000)	Community adults (*230*) Age$_M$ = 62.6 (12.5)	X	-.01	Schwall (2008)	Community adults (*N* = 387) Age$_M$ = 52 (7.2) [>40]	X	-.18**
Noone et al (2010)[a]	Community adults (*N* = 1,532) Age$_M$ = 53.4 (2.89) [49-60]	X	.00	Schwall (2008)	Community adults (*N* = 331) Age$_M$ = 47.2 (10.6)	X	-.36**
Hershey et al. (2010)	Community adults[f] (*N* = 975) Age$_M$ = 42.5 (10.5) [25-65]	X	.07	Schwall (2008)	Community adults (*N* = 368) Age$_M$ = 48.6 (10.3)	X	-.19**
				Kessler & Staudinger (2011)	Community adults (*N* = 277) Age$_M$ = 47.5 (16.8)	X	-.67**
				Lang & Carstensen (2002)	Community adults (*N* = 480) Age$_M$ = 55.7 (5.8)	X	-.70**
				Treadway et al. (2012)	Full-time adults (*N* = 291) Age$_M$ = 30.6 (10.5)	X	-.35**
				Kooij (2010)	University workers (*N* = 662) Age$_M$ = 44.2 (10.9)	X[b]	-.67**
				Bluck & Alea (2009)	Community adults (*N* = 185) Age$_{M(Young)}$ = 19.31 (2.80) Age$_{M(Old)}$ = 73.04, (7.53)	X[e]	-.63**
				Löckenhoff, O'Donoghue, & Dunning (2011)	Community adults (*N* = 98) Age$_M$ = 52.00 (20.5)	X	-.43**
Retirement Planning							

Table 3 (continued)

Study	Sample (*N*)	Design	ES *r*	Study	Sample (*N*)	Design	ES *r*
Hershey et al. (2007)	Full-time employees (*265*) Age_M = 36.3 (6.18) [25-45]	X	.41**	Schwall (2008)	Community adults (*N* = 331) Age_M = 47.2 (10.6)	X	.32**
Hershey & Mowen (2000)[a]	Community adults (*230*) Age_M = 62.6 (12.5)	X	.50**				
Noone et al (2010)[a, d,]	Community adults (*N* = 1,532) Age_M = 53.4 (2.89) [49-60]	X	.10 - .37**				
Jacobs-Lawson & Hershey (2005)[a]	Full-time employees (*N* = 265) Age_M = 36.2 (6.18) [25-45]	X	.26**				
Intended Retirement							
Study	**Sample (*N*)**	**Design**	**ES *r***	**Study**	**Sample (*N*)**	**Design**	**ES *r***
Noone et al (2010)[a, c]	Community adults (*N* = 1,532) Age_M = 53.4 (2.89) [49-60]	X	-.25**	Schwall (2008)	Community adults (*N* = 233) Age_M = 47 (10.6)	X	.26**
				Schwall (2008)	Community adults (*N* = 387) Age_M = 52 (7.2) [>40]	X	.11*
				Schwall (2008)	Community adults > 40 (*N* = 331) Age_M = 47.2 (10.6)	X	.28**
				Schwall (2008)	Community adults > 40 (*N* = 368) Age_M = 48.6 (10.3)	X	.09*
				Kooij (2010)	University workers (*N* = 662) Age_M = 44.2 (10.9)	X[b]	-.04

Note. Standard deviations in parentheses, age inclusion criteria inside brackets. [a] originally conceptualized instrument, [b] cross-sectional analysis of longitudinal data, [c] age-adjusted, [d] range across four domains, [e] extreme-groups, [f] cross-national sample. X = cross-sectional. * $p < .05$, **, $p < .01$.

CHAPTER 5

RETROSPECTIVE-OBSERVATION STUDY

5.1 Overview

The retrospective observation study adopted two latent variable-analytic frameworks to examine two sources of validity evidence. Specifically, IRT analyses were conducted to examine the psychometric properties of the C-FTP scale and items. In addition, SEM was utilized to examine the proposed structural model, primarily to inform the subsequent experiment. Before presenting formal hypothesis testing, for each analytic framework, a separate Methods section will document the data treatment protocol[27] and auxiliary analyses. The IRT subsection is presented first and SEM hypothesis testing will conclude the chapter. Generally, the informational format follows reporting standards of the *Archives of Scientific Psychology*.

5.2 Retro-Obs Methods (IRT)

Participants identified as missing all datum (100% incomplete, i.e., system-missing, or, unit nonresponse) across the focal variable future time perspective (FTP) were removed from the dataset, resulting in N = 4,231 participants. Further exclusion criteria was substantively motivated in order to permit stronger inferences of the FTP instrument's psychometric properties to the target population of interest, that is, to normal community-dwelling older adults. Two age-bands with fixed upper-bounds (<

[27] Data treatment is a variant of data processing in information systems nomenclature, that is, "the collection and manipulation of items of data to produce meaningful information" (French, 1996), but it is held distinct from "mining". The author also purposefully retains the term 'treatment' to connote the inherent subjective investigator judgment and decision-making in the handling process. In the current data, treatment encompasses demographic and quality validation, sorting, aggregation, and missing value pattern analysis, prior to formal hypothesis testing and auxiliary analyses.

80-years) and lower-bounds at 60- and 65-years were selected in order to approximate equitable distribution of public pension-eligibles (U.S. Social Security Agency, 2014).

Because the 'job status' item response format was additive, i.e., multiple answer, joint-retiree and -worker respondents were excluded in order to reduce cross-classification contamination (*N* = 47). Figure 1 summarizes the stratified samples. Additional summary descriptives pertinent to the data editing procedure (reductions) is provided in Table 1.

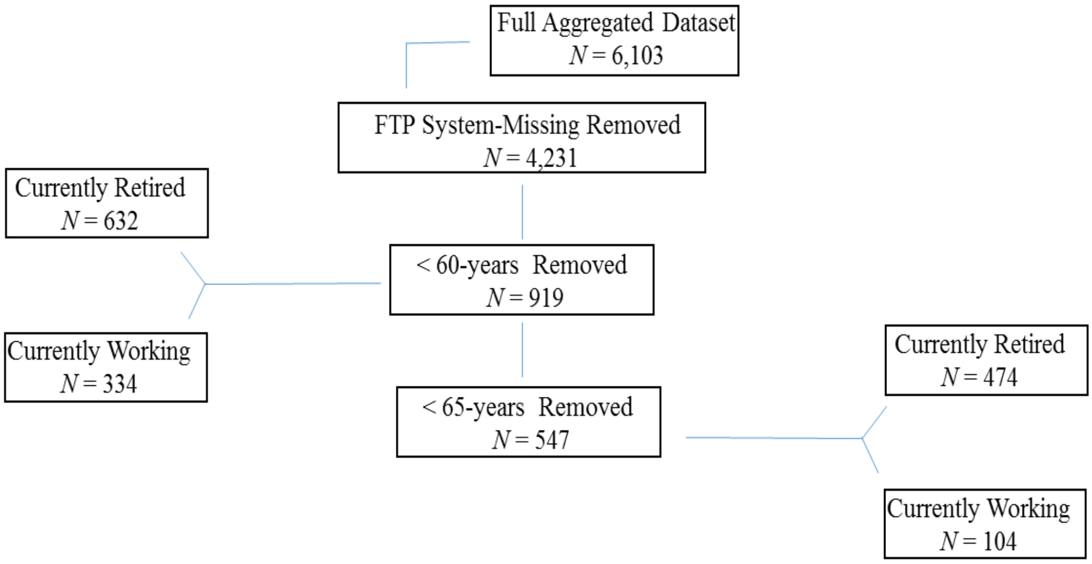


Figure 2. Dendogram summary of stratified subsamples for IRT analyses.

Summary descriptive statistics reported in Table 1 illustrate sample characteristics across the hierarchical strati. Demographic proportions across strati reflect historic population estimates, while relative within-stratum proportions appear approximately equivalent. For example, gender and race proportions change markedly across adult and older-adult populations, while neither change drastically within the older adult population. The data provides limited confidence in population inferences from obtained test findings in the sample data, also, from the panel's sampling frames developed on current population census data (U.S. Censor Bureau, 2014).

Table 4

Summary Sample Characteristics over Hierarchical Strati.

Sub-Samples	Age $\bar{x}$	Sex F$\hat{p}$ / M $\hat{p}$	Race Wh$\hat{p}$ / Bl$\hat{p}$ / As$\hat{p}$ / Ot$\hat{p}$
Full Aggregate$_{(N=6,103)}$	46.07	.59 / .41	.71 / .11 / .02 / .09
FTP-Completes$_{(N=4,231)}$	47.14	.60 / .40	.79 / .11 / .02 / .07
<60-Years Remov$_{(N=919)}$	66.95	.52 / .48	.86 / .06 / .01 / .02
<65-Years Remov$_{(N=547)}$	70.79	.51 / .49	.89 / .05 / .01 / .05

Note. 1. SLE $\hat{p}_{\text{missing}}$ relative to <75-years age for item relevance (probability to live to age 75). Workers and retirees Ì job status ∴ Σ < 1 . SLE = subjective life expectancy, Wh = White, Bl = Black, As = Asian, Ot = Other.

In addition to demographic sample characteristics for population inference, information regarding nonresponse patterns is provided in Table 2. Specifically, information is provided on relative proportions of workers and retirees affected by data reductions, statistical tests of user-missing data on the focal questionnaire instrument (FTP) across reductions, as well as gross-referent indication of potential reactivity (response refusal) based on descriptive summaries of the subjective life expectancy (SLE) item across data reductions. These findings are addressed below.

First, the relative proportions of workers to retirees was affected by the data reductions. Second, in order to provide statistical inference regarding the nature of the expected change[28] vis-à-vis potential nonresponse bias, a series of Little's (1988) missing-completely-at-random (MCAR) chi-square tests was conducted on the FTP

[28] Composition of older participants expected to shift relative proportion from workers to retirees.

questionnaire. As indicated, item non-response to the FTP instrument was considered missing-completely-at-random. This finding has implications for the generalizability of inferences from statistical test, in spite of the expected change in proportions of sampling as a function of data reductions. That is, the underlying mechanism of the response propensity, i.e., second-phase sampling, may be considered statistically unrelated to the composition of the sample in terms of workers and retirees. Third, summary descriptives across the SLE item indicated comparable values and response probabilities.

In summary, given the negligible amount of missing data on the focal-FTP questionnaire < 1%), as well as evidence for the user-missing data to be MCAR, the IRT data analyses proceeded with complete-case analysis (i.e., listwise deletion). Uunder MCAR assumptions, complete-case analysis reduces power, but yields unbiased parameter estimates. A more thorough elaboration of missing data is also reserved for the SEM analyses subsection.[29]

[29] It should be noted that IRT approaches are widely available for the modeling of ignorable (Finch, 2008) and nonignorable missing data mechanisms (see, Allison, 2001; DeMars, 2002; von Davier, & Xu, 2010).

Table 5

Summary Job Status Proportions, MCAR-Tests of FTP Instrument, and SLE Summary Descriptives over Sample Strati.

Sub-Samples	$\hat{p}_{\text{workers}}$	$\hat{p}_{\text{retirees}}$	Little's MCAR-$X^2_{\text{(All FTP)}}$	SLE $\bar{x}$ (s) [$\hat{p}_{\text{missing}}$][1]
Full Aggregate$_{(N=6,103)}$	.415	.124	X^2 (164, 6,103) = 193.41, p = .06	.449 (.19)
FTP-Completes$_{(N=4,231)}$	59.5	17.8	X^2 (148, 4,231) = 155.47, p = .32	.60 (.45)
<60-Years Remov$_{(N=919)}$	.363	.688	X^2 (22, 919) = 13.08, p = .93	.65 (1.13)
<65-Years Remov$_{(N=547)}$	.198	.867	X^2 (11, 547) = 3.88, p = 97.	.63 (2.14)

Note. 1. SLE $\hat{p}_{\text{missing}}$ relative to <75-years age for item relevance (probability to live to age 75). Workers and retirees Ì job status ∴ Σ < 1, SLE = subjective life expectancy, Remov = removed.

5.3 Retro-Obs Results (IRT)

The items of the Future Timer Perspective (FTP) questionnaire were calibrated and parameterized using the generalized-partial credit model (GPCM; Masters, 1982; Muraki, 1990; 1992). The parametric model was initially selected for its balance to identifiability of measurement properties and flexibility-fit for polytomous data (Thissen and Steinberg, 1986). Specifically, the model assumes that adjacent response-category logits are nondecreasing functions of the latent trait, θ. It is a generalized form of Master's (1982) adapted Rasch model for polytomous data (1PL / *k* categories for each item, *j*), the partial credit model, by relaxing the constant discrimination assumption across items. Consequently, the GPCM model is an adapted version of the 2PL model for polytomous data, expressed

as, $P_{ix}(\theta) = \dfrac{EXP\left[\sum_{j=0}^{k} \alpha\left(\theta - \delta_{ij}\right)\right]}{\sum_{r=0}^{m_i}\left[EXP \sum_{j=0}^{k} \alpha\left(\theta - \delta_{ij}\right)\right]}$ for the probability of responding in a given

category *k* for each item, *j*, and across all *j* items as,

$$P_{ix}(\theta) = \frac{EXP\left[\alpha(k(\theta - \beta_i) + \sum_{j=1}^{k} d_k)\right]}{\sum_{r=0}^{m_i}\left[EXP(\alpha(k(\theta - \beta_i) + \sum_{j=1}^{k} d_k)\right]}$$

, where k_i = the number of response categories, δ_{ij} = the step difficulty parameter (within-item). The step difficulty parameter can be further decomposed into $\delta_{ij} = \beta_i + \sum_{j=1}^{k} d_k$, where *d* represents a deviation index for each adjacent category from the first category, which is fixed at 0 for identification. As a result, the category-means may vary about 0 as a function of the overall item location (difficulty) parameter.

Because the discrimination parameter α is permitted to vary across items, but remains constant across score categories within-item, each item will have a total of *k* parameter estimates per item, a *b* estimate for every step function (k – 1) plus a discrimination parameter (α). A given FTP item measured on the 7-point likert-type scale, therefore, will have six location estimates (category deviations as a function of item location) and one discrimination estimate. Across the 12-item instrument, a total of 7 x (12) = 84 parameters was estimated.

5.3.1 Comparative Model Fit

The theoretical suitability of the GPCM model for the current FTP data was tested, specifically, against another polytomous parametric model, the partial credit model (PCM, 1-PL analogue for polytomous data). A model-comparison was conducted in the largest subsample (*N* = 917), with an initial assessment of differential item functioning across workers and retirees. A combination of statistical tests (M_2), residual-based indices (RMSEA), and information criteria (log-likelihood ratios, AIC, BIC) were evaluated for comparing GPCM and PCM model fit. Bock-Aitken estimation is employed for both models. Results are presented in Table 3 and supported the retention of the GPCM model as a relatively better fit to the FTP sample dataset.

Table 6

Model Fit Indices for Models using the FTP Questionnaire.

Model	[1]M_2 (df), p-value	-2lnL	AIC	BIC	RMSEA
PCM	8921.92 (2108), *p* < .00	35425.55	35597.55	36007.54	.06
GPCM	7467.38 (2086), *p* < .00	35175.71	35515.71	36326.15	.05

Note. 1. Response categories were reduced from 7 to 5 for local identifiability in order to permit estimation of the M_2 statistic, only (Cai & Hansen, 2013). -2lnL = -2 log likelihood, AIC = Akiake information criterion, BIC = Bayesian information criterion, PCM = partial credit model, GPCM = Generalized-PCM.

5.3.2 Retro-Obs IRT Hypothesis Tests

In order to test the measurement invariance of C-FTP across employment status categories, pairwise comparisons of compact and augmented models were conducted for each item following the procedure from Kim, Kim, and Kamphaus (2005). The stepwise procedure was tractable for relatively the small number of items comprising the FTP instrument (j = 12). Statistics for model-comparisons are indexed as likelihood-ratio difference indicators (G^2) between -2 log likelihoods of the full measurement-invariance model (compact) and pairwise-item partial invariance models (augmented). It may be expressed as,

$$G^2 = -2\ln L_{(\text{full measure-invaraiance})} - (-2\ln L_{(\text{partial measure-invariance})})$$

The analysis provided more comprehensive information regarding measurement noninvariance of the FTP instrument across workers and retirees. Results are presented in Table 4, below. For statistically significant G^2 indices, analyses were repeated with separate equality constraints imposed on item discrimination and location parameters to identify plausible sources of misfit. Formal hypotheses are also restated below.

R-H1: The location parameters of the FTP items will be higher among retirees, relative to employees (i.e., difficulty parameters [$b_{.j}$]).

R-H2: The discrimination parameters ($a_{.j}$) of the FTP items will be greater among retirees, relative to employees.

Initial inspection of model results reported in Table 6 indicated preliminary evidence of DIF on select-FTP items across workers and retiree samples. Specifically, approximately half of C-FTP's twelve items indicated potential DIF, however, the compact – augmented model comparisons do not permit identification of the utility of individual items, that is, their relative information across groups. In addition, initial descriptive statistics from the compact – augmented model comparisons indicated lower mean scores and greater variance across all items, save one, for retirees relative to workers.

Table 7

Stepwise Compact-Augmented Model Comparisons of Future Time Perspective Item Measurement Invariance across Workers and Retirees.

		Reference(Workers)								Focal(Retirees)									
Item	G^2	α	b_1	d_2	d_3	d_4	d_5	d_6	d_7	α	b_1	d_2	d_3	d_4	d_5	*d6*	*d7*	$\mu_{(Ret)}$	$\sigma_{(Ret)}$
1	244.09*	2.50	-0.36	1.30	0.64	0.37	-0.33	-0.76	-1.22	2.06	-0.18	1.09	0.36	0.29	-0.37	-0.79	-0.58	0.65	-0.18
2	20.72*	0.77	-0.41	0.97	0.77	0.30	-0.46	-0.07	-1.51	0.38	-0.39	0.74	0.93	0.09	0.19	-0.53	-1.42	-0.19	0.67
3	7.37	2.09	-0.21	1.43	0.55	0.16	-0.44	-0.76	-0.94	2.49	-0.20	1.04	0.40	0.28	-0.30	-0.66	-0.76	-0.18	0.64
4	53.33*	5.82	1.98	-0.28	0.26	0.00	0.96	1.98	0.47	2.48	-0.38	1.37	0.55	0.15	-0.26	-0.85	-0.95	-0.25	0.88
5	29.92*	0.82	0.54	1.41	0.22	-0.11	-0.34	-0.77	-0.41	1.29	0.37	0.93	0.17	0.06	-0.36	-0.70	-0.11	-0.18	0.63
6	21.55*	0.81	0.39	1.21	0.19	0.56	-0.92	-0.95	-0.10	1.03	0.16	0.61	0.39	0.17	-0.34	-0.68	-0.16	-0.19	0.65
7	7.36	0.77	-0.41	0.97	0.77	0.30	-0.46	-0.07	-1.51	0.38	-0.39	0.74	0.93	0.09	0.19	-0.53	-1.42	-0.19	0.67
8	29.92*	1.32	-0.02	1.53	0.12	0.32	-0.61	-0.41	-0.95	0.72	-0.22	1.61	0.19	0.35	-0.71	-1.04	-0.40	-0.20	0.66
9	12.30	1.23	0.01	2.02	0.36	-0.04	-0.47	-0.71	-1.16	1.10	-0.07	1.68	0.16	0.22	-0.32	-0.98	-0.77	-0.19	0.66
10	4.99	0.24	-0.30	2.38	0.54	-0.03	-1.13	-1.07	-0.70	0.29	-0.37	2.43	0.04	1.03	-0.89	-1.13	-1.49	-0.19	0.66
11	9.42	0.16	1.21	4.10	0.18	-1.61	1.75	-0.93	-3.48	0.28	0.43	1.87	1.04	-0.44	-1.10	-0.91	-0.46	-0.19	0.66
12	8.43	-0.02	-17.28	-33.22	-4.13	1.55	14.54	-7.19	28.44	0.16	2.04	2.66	1.37	1.34	-2.44	-0.04	-2.90	-0.19	0.66

Note. * = two-tailed asymptotic-$X^2_{(7)}$ critical value = 18.48 with α = .01. N = 869. All G^2 values tested for significance with 7 degrees of freedom (*df*).

Following compact-augmented model comparisons, formal DIF analyses were conducted in order to test for noninvariance of FTP items across workers and retirees. The formal DIF analyses has the added advantage of partitioning variance-sources of DIF into slope and location parameter estimates within a single estimation, thereby, improving type-1 error rate control. The DIF analyses proceeded with three anchoring methods in order of statistical stringency: First, all items served as anchors across groups. Second, all items indicating invariance from the compact-augmented model comparisons served as anchors across groups (all-other).[30] Third, items indicating measurement invariance from the first all-anchor procedure served as anchors across groups.

Table 8

Summary DIF Statistics by Slope and Location Parameter Estimates for Anchor-All-Items

Item	$X^2_{(location)}$	*Df*	*p*-value	$X^2_{(slope)}$	*Df*	*p*-value
1	22.9	6	.00	1.2	1	.27
2	10.6	6	.10	2.80	1	.09
3	25.5	6	.00	.20	1	.69
4	4.8	6	.58	.20	1	.69
5	27.5	6	.00	.20	1	.62
6	21.7	6	.00	.20	1	.64
7	6.7	6	.35	.30	1	.59
8	13.3	6	.04	.20	1	.69
9	14.1	6	.03	.00	1	.87
10	5.2	6	.53	.50	1	.48
11	13.9	6	.03	.80	1	.38
12	9.5	6	.15	2.40	1	.12

Note. N = 869.

[30] This may also be known as a form of non-compensatory DIF.

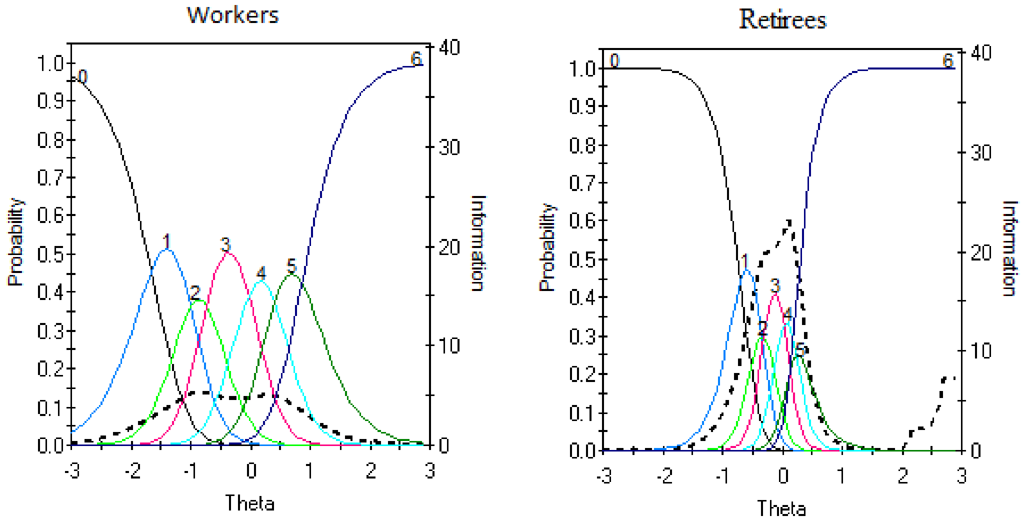
Table 9

Summary DIF Statistics by Slope and Location Parameter Estimates for Anchor-All-Other Items

Item	$X^2_{(location)}$	*Df*	*p*-value	$X^2_{(slope)}$	*Df*	*p*-value
1	22.7	6	.00	1.00	1	.32
2	7.1	6	.32	8.5	1	.00
3	20.1	6	.00	0.10	1	.72
5	5.0	6	.55	15.6	1	.00
6	19.3	6	.00	6.90	1	.01
8	10.5	6	.11	6.90	1	.01
9	12.2	6	.06	0.10	1	.77
11	8.0	6	.24	4.00	1	.04

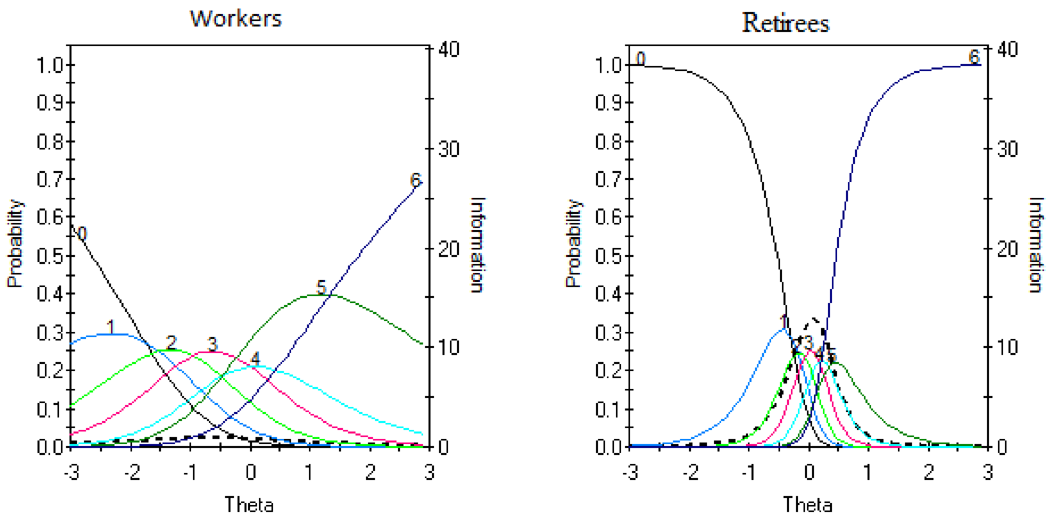
Note. *N* = 869.

In addition to the statistical analyses testing for DIF, combined-trace lines and item characteristic curves (ICCs) from the ‘anchor-all’ DIF procedure for all FTP items across workers and retirees is displayed in Figures 3 – 13 below. Item trace line models the conditional probability of a correct response across levels of theta for which the item provides information (Lord, 1977). Comparing the trace lines between groups for the ‘anchor-all’ model provides a liberal illustration of potential DIF, also, information on item characteristics within each group. Item content is centered below the side – side comparison graphs.



FTP 1 - "Many opportunities await me in the future."

Figure 3. Combined-Trace Lines and ICCs for the Ancor-All-Items Model across Workers and Retirees for Item 1.



FTP 2R - "There are only limited possibilities in my future."

Figure 4. Combined-Trace Lines and ICCs for the Ancor-All-Items Model across Workers and Retirees for Item 2.

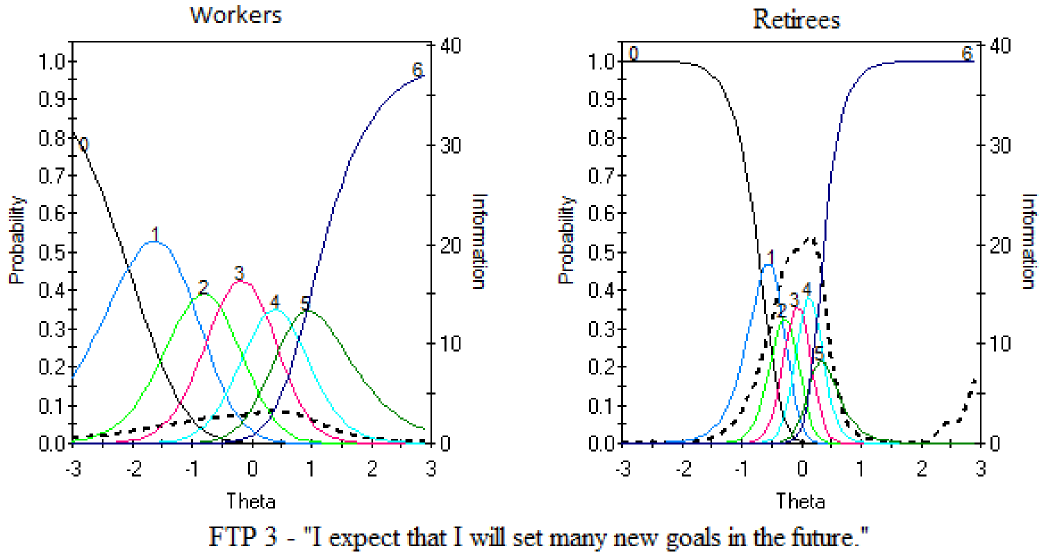


Figure 5. Combined-Trace Lines and ICCs for the Ancor-All-Items Model across Workers and Retirees for Item 3.

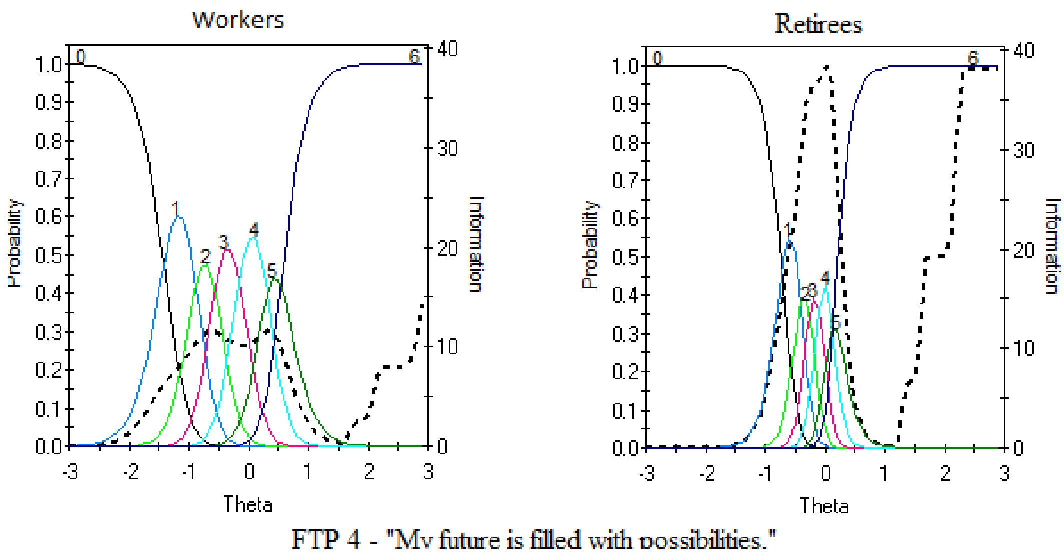


Figure 6. Combined-Trace Lines and ICCs for the Ancor-All-Items Model across Workers and Retirees for Item 4.

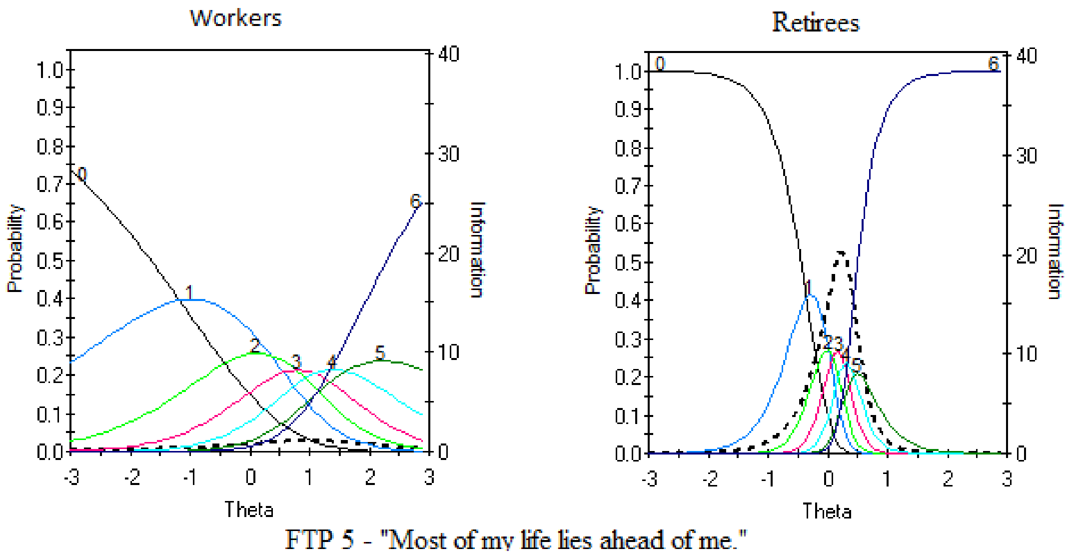


Figure 7. Combined-Trace Lines and ICCs for the Ancor-All-Items Model across Workers and Retirees for Item 5.

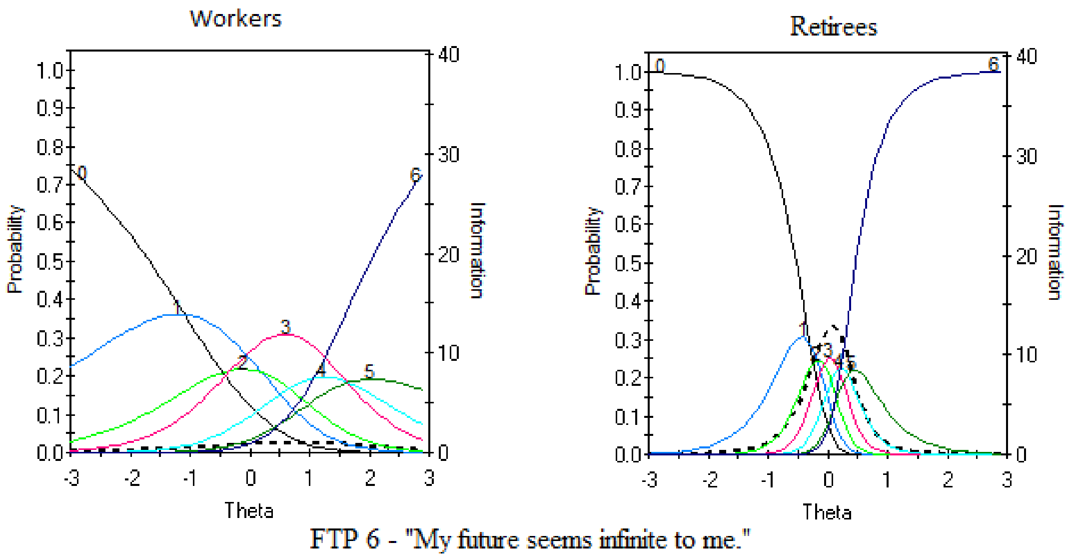
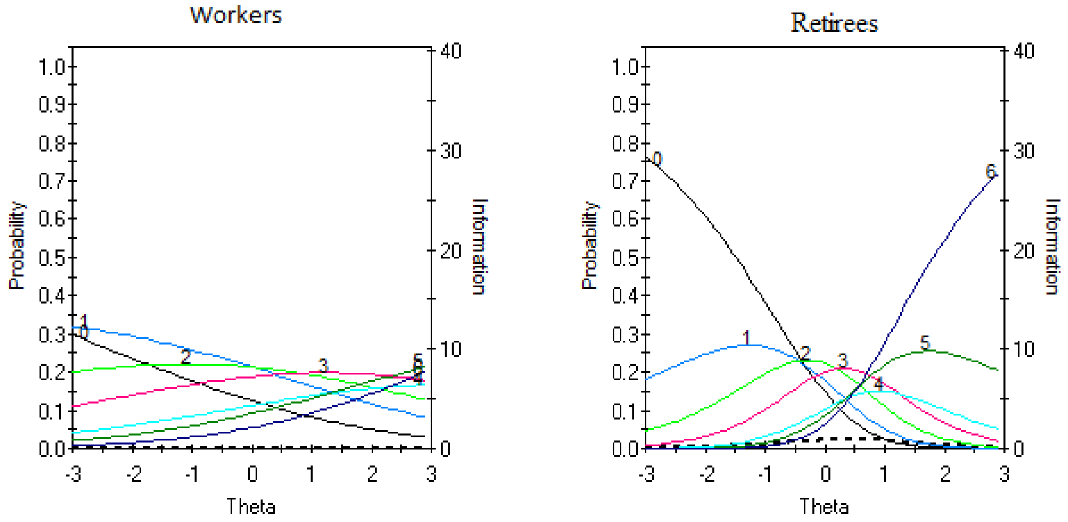
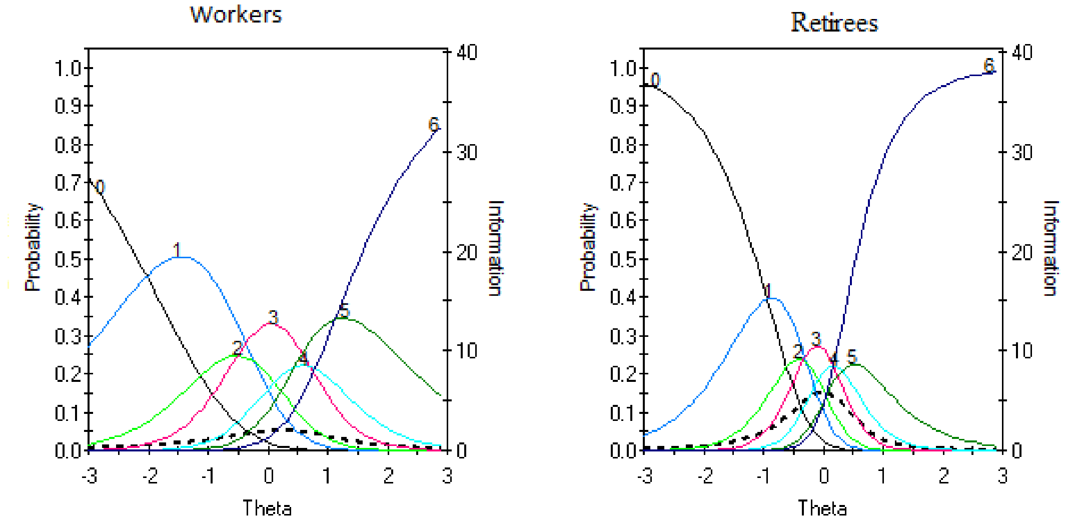


Figure 8. *Figure 7*. Combined-Trace Lines and ICCs for the Ancor-All-Items Model across Workers and Retirees for Item 6.



FTP 7R - "I have limited time left to live my life."

Figure 9. Combined-Trace Lines and ICCs for the Ancor-All-Items Model across Workers and Retirees for Item 7.



FTP 8 - "I could do anything I want in the future."

Figure 10. Combined-Trace Lines and ICCs for the Ancor-All-Items Model across Workers and Retirees for Item 8.

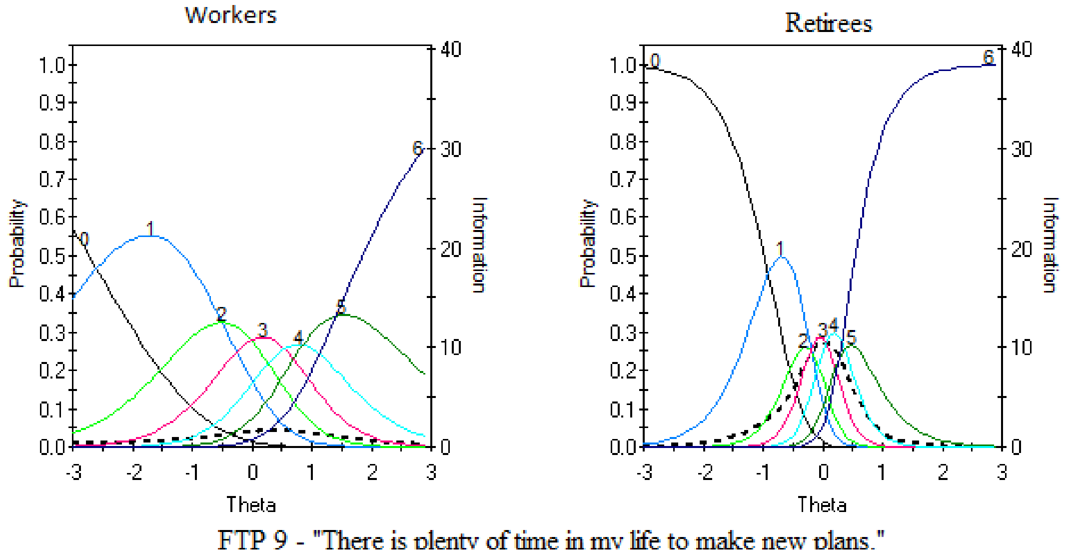


Figure 11. Combined-Trace Lines and ICCs for the Ancor-All-Items Model across Workers and Retirees for Item 9.

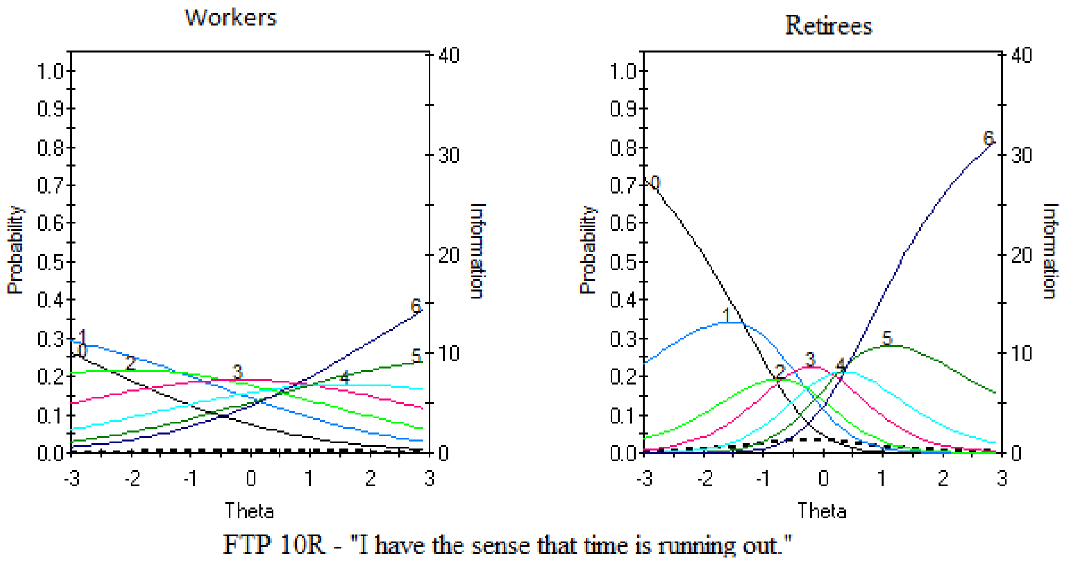


Figure 12. Combined-Trace Lines and ICCs for the Ancor-All-Items Model across Workers and Retirees for Item 10.

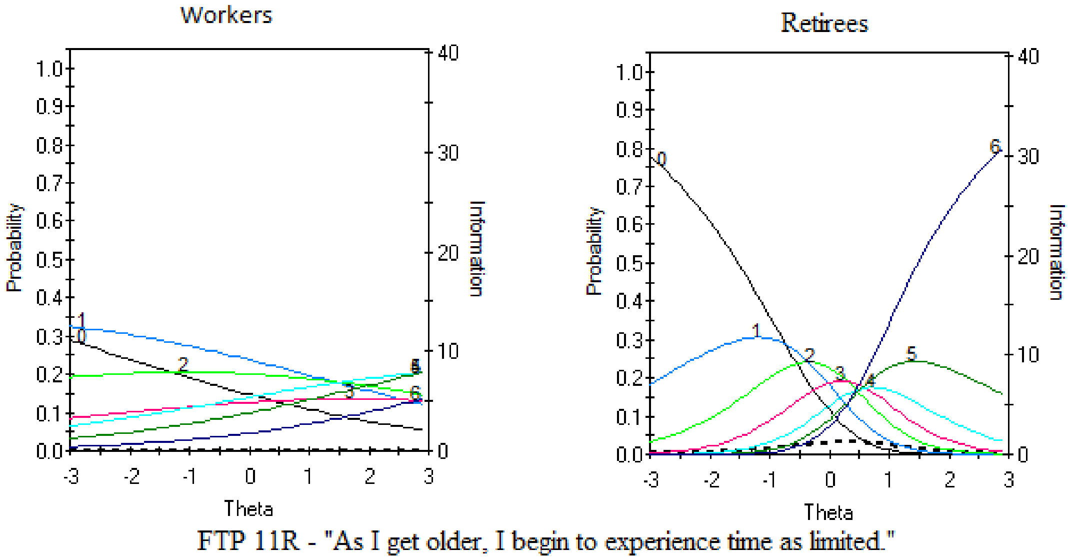


Figure 13. Combined-Trace Lines and ICCs for the Ancor-All-Items Model across Workers and Retirees for Item 11.

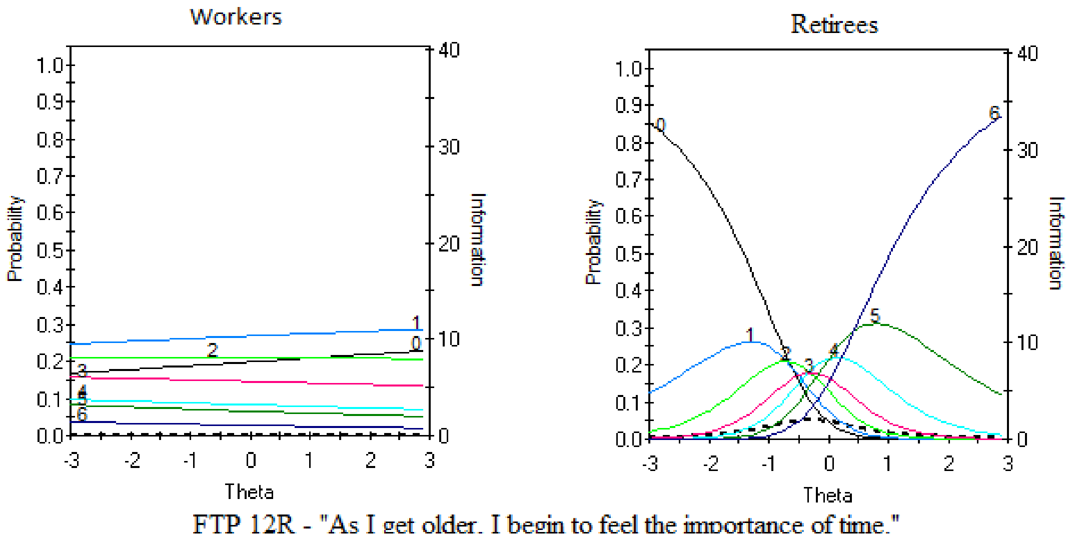


Figure 14. Combined-Trace Lines and ICCs for the Ancor-All-Items Model across Workers and Retirees for Item 11.

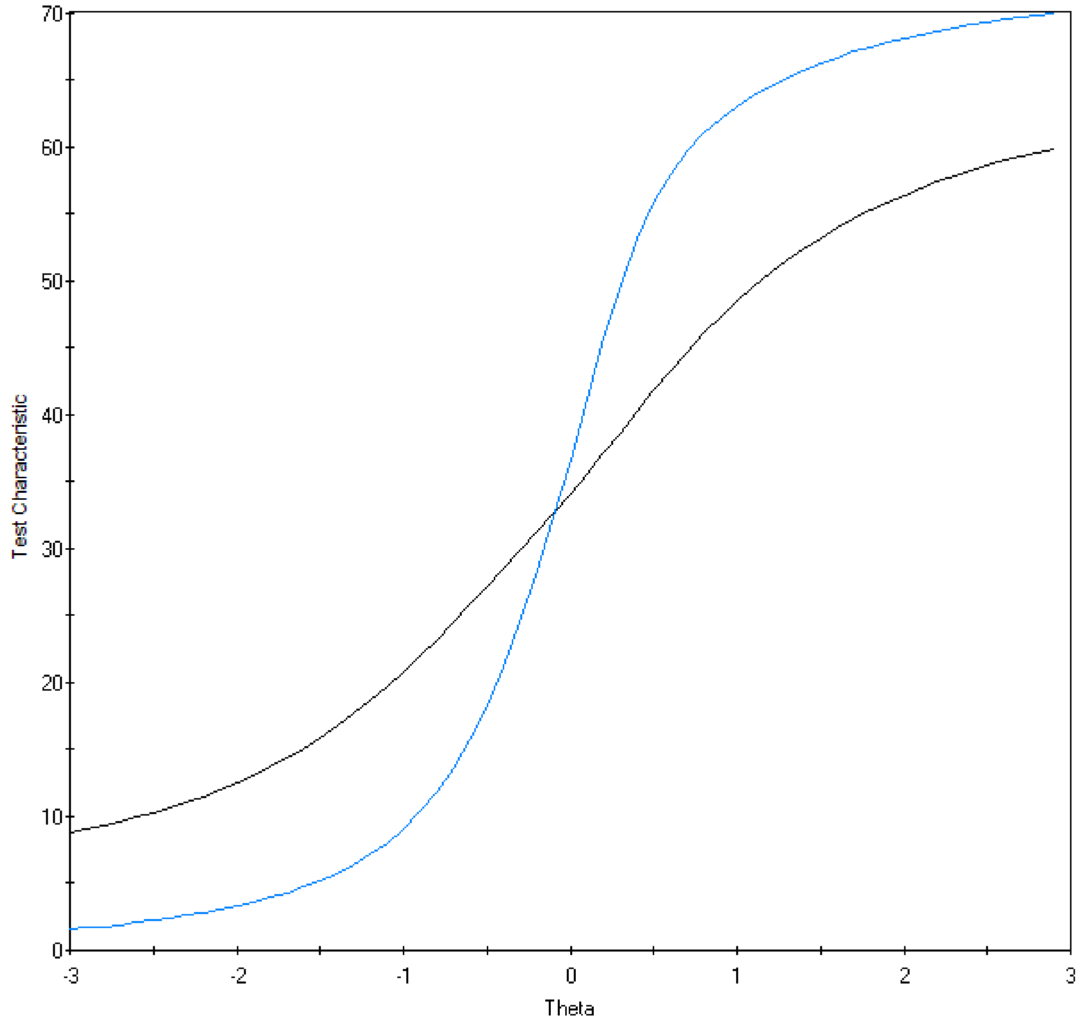


Figure 14. Combined FTP Test-Characteristics Curve for Workers and Retirees, Anchoring on All Items. *Note.* = *Workers*, = Retirees

Table 10

Item Information Function Values by Workers and Retirees at Five Levels of θ.

Item	Sample	θ = -2	θ = -1.2	θ = 0	θ = 1.2	θ = 2
1	Workers	1.75	4.52	4.77	2.02	0.35
	Retirees	0.79	2.48	5.02	1.36	0.19
2	Workers	0.56	0.83	0.79	0.39	0.22
	Retirees	0.31	0.44	0.46	0.28	0.18
3	Workers	1.23	2.04	2.91	1.84	0.49
	Retirees	0.68	2.00	4.53	1.90	0.26
4	Workers	1.36	7.74	10.21	1.05	3.22
	Retirees	0.25	0.70	3.04	2.62	0.48
5	Workers	0.20	0.33	0.74	1.05	0.68
	Retirees	0.25	0.70	3.04	2.62	0.48
6	Workers	0.24	0.54	0.81	0.95	0.62
	Retirees	0.27	0.79	2.90	1.40	0.36
7	Workers	0.05	0.06	0.07	0.07	0.07
	Retirees	0.11	0.15	0.20	0.22	0.20
8	Workers	0.50	0.91	1.90	1.18	0.48
	Retirees	0.42	0.77	1.39	0.80	0.33
9	Workers	0.47	0.72	1.50	1.24	0.60
	Retirees	0.54	1.04	2.50	1.37	0.45
10	Workers	0.09	0.10	0.11	0.10	0.09
	Retirees	0.21	0.28	0.31	0.24	0.18
11	Workers	0.03	0.04	0.04	0.05	0.05
	Retirees	0.14	0.19	0.28	0.30	0.25
12	Workers	0	0	0	0	0
	Retirees	0.06	0.08	0.10	0.11	0.12
	Workers	7.49	18.70	24.86	10.93	7.87
Test Info		(.37)	(.23)	(.20)	(.30)	(.36)
	Retirees	7.12	15.43	32.08	13.84	5.35
		(.37)	(.25)	(.18)	(.27)	(.43)

Note. *N* = 869. All-items-anchor method. Expected standard errors in paranetheses.

5.3.3 Summary of Findings

Compact – augmented model comparison provided preliminary evidence for differential item functioning (DIF) across older workers and retirees. Formal DIF analyses with various anchoring methods indicated item noninvariance (inequivalence). An inspection of the comparative location parameter estimates supported supported hypothesis 1 regarding higher location parameters of FTP among retirees, relative to older workers. This is also illustrated by the right-shifted item information functions for the retiree sample, relative to older workers (Figures 3 – 14). In addition, evidence supported the second hypothesis regarding higher discrimination parameter estimates among retirees, relative to older employees. This is also illustrated in Figures 3 – 14 by the relative height of the item characteristics curves. Finally, the magnitude of the instrument-wide discrimination differences of retirees versus workers is depicted in Figure 14.

Extrapolating from the estimated item information across levels of theta, in may be inferred that Carstensen's FTP instrument provides little information on FTP among older workers, and more information among retirees, particularly at above-average levels of the underlying latent trait. The next section introduces the SEM Methods and Results for testing a structural model of FTP and intended retirement.

5.4 (SEM) Methods

5.4.1 Data Treatment

Treatment of data was conducted in SPSS v22.0. Merging of all data files with candidate variables resulted in an aggregated dataset of N = 6,103 participants. A dendogram overviewing the stratified subsamples is presented below.

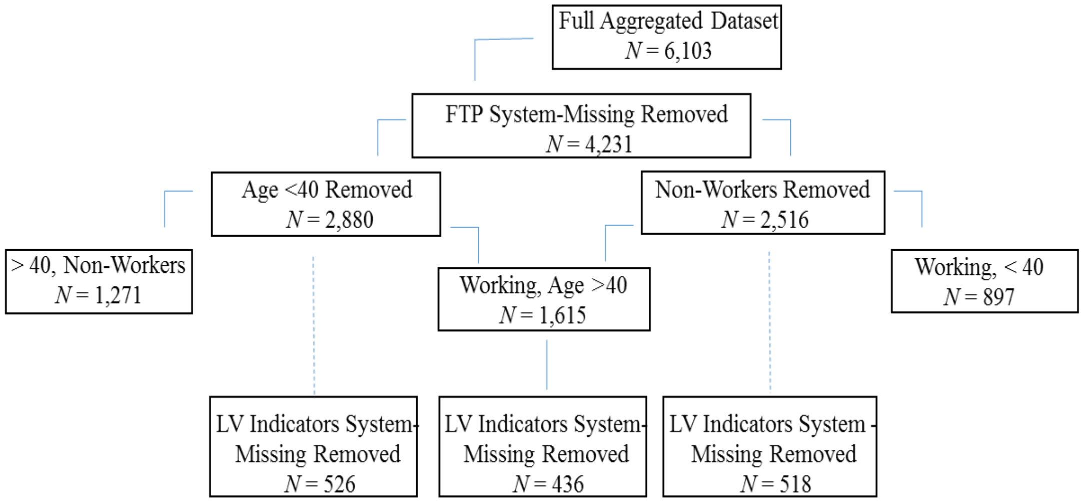


Figure 15. Dendogram summary of stratified subsamples for SEM analyses.

Duplicating the IRT method, participants identified as missing all data (100%, i.e., system-missing, or, unit nonresponse) across the focal variable future time perspective (FTP) were removed from the dataset, resulting in N = 4,231 participants.[31] The target-age population (> 40-years) for analyses was identified with a computed age variable informed by combined demographic information across all constituent data files, resulting in age-identification of 96.5% of participants in the dataset. Remaining unidentified participants' ages (N = 147) were manually inputted

[31] For a recent, in-depth review of nonresponse error in questionnaire data from the Information Systems (IS) field, interested readers should see Sivo, Saunders, Chang, & Jiang (2006).

from interpolated estimates of demographic info on constituent data files and dated-indicators (para-data) of the surveys informing computation of the age variable (edit).

From these data, participants < 40-years of age were excluded, resulting in *N* = 2,825 participants. This age cut-point was determined by three factors in order of consideration, 1) procedurally, approximation of the proposed age-range for the subsequent experimental studies, 2) empirically, meta-analytic findings indicating developmental trends of retirement ideation (Topa, Moriano, Depolo, Alcover, & Morales, 2009), and 3) operationally, the protected cut-age identified in the Age Discrimination in Employment Act (ADEA; 1967). It is also used as a lower-bound from which refined age-bands may be utilized for comparative analyses.[32]

From these data, 'non-workers'[33] were removed, resulting in *N* = 1,619. This decision was guided by the retirement intention outcome of interest in the current structural model, specifically, as a relevant and meaningful percept for respondents. The '> 40-year old workers' subsample was then cross-checked for system-missingness on all latent factor indicators in the proposed measurement model. This cross-check led to a final, base sample of *N* = 436. Table 1 summarizes the patterns of system-missing data over the subsamples, however, it should be noted that the figures are gross upper-bounds primarily for data structure-illustration purposes. A refined approach is detailed in a later subsection on missing data.

[32] It should be noted, because this lower-bound was technically lower than the inclusion criteria for the proposed experiment (age 49 – 65-years), an approximated upper-bound was also imposed to band parameter estimates (> 80-years, *N* = 47).
[33] Non-workers were self-identified as either unemployed, on-leave, disabled, retired, housemaker, or 'other'.

Table 11

Summary Statistics for Pattern Analyses of System-Missing Data across Datasets

Variable (*J* / *K*)	Full (*N* = 6,103) % system-missing[1]	FTP-Removed (*N* = 4,231) % system-missing	>40, Work (*N* = 1,619) % system-missing
Intended Retirement (1)	78.2	69.5	60.1
SLE (1)	17.5	16.0	8.2
FTP (6)	30.7	-	-
Optimism (6)	52.4	37.2	29.7
Choice Deferral (5)	82.5	74.9	69.1
Impulsivity (5)	82.5	74.9	69.1
Sunk Cost (4)	82.5	74.9	69.1

Note. *J* / *K* = # of measured indicators / latent factor (indicated in parentheses). 1. System-missing defined by 100% missing measured indicators, within factor.

5.4.1.1 Quality Check. An ad-hoc data quality check was implemented in order to screen for potential careless responders. The procedure comprised computations of two consistency pattern indices (Meade & Craig, 2012). All auto-identified participants were subsequently reviewed manually across constituent data points for potential exclusion. Only observably consistent careless responders were excluded, also, after cross-checking other potential sources of observation error, limited to ‘data entry’ for archived data in the present case (Anscombe, 1960). The first computed index was a consistency dummy-variable on the focal FTP measure, which comprised a near-balance of reverse-scored items ($j_{reverse} / j_{total}$ = 5 / 12). One participant was identified and excluded upon review. A second consistency index was more liberal in identifying all pairwise responders across joint-standard and -reverse-scored items who responded in the furthest two- extreme categories. One participant was identified and excluded upon review, resulting in a quality-checked, base sample of *N* = 434.

5.4.1.2 Missing data. Missing data handling is nearly requisite to education and social sciences research (Roth, 1994, Dillman, 1989). In the current retrospective observation study, in addition to analyses on variant complete- and available-case datasets, Rubin’s (1987) recommendation[34] for the routine generation of multiple datasets under different missing data models and assumptions is implemented. This practice permits estimation of missing data handling error, more generally (c.f., Cronbach’s ‘units’ of observation; also, Fallegi & Holt, 1976).

[34] Rubin (1987) technically prescribed multiple imputed datasets for estimating imputation error, but this is generalized with relaxed assumptions for comparing non-imputation analytic results, as well. Fallegi and Holt (1976) proposed a systematic approach aligning ‘automatic’ (computerized) edits and imputations.

Missing value pattern analyses. Participants identified as careless responders did not have missing data, and so were also excluded from the missing data analysis. Two preliminary, ordered steps to understanding the missing data pattern included, 1) The precision of its composition, specifically, differentiating non-responders from potential undercoverage (whether due to undersampling or unit- nonresponse, contrast, item non-response),[35] and 2) The accuracy of retained sample characteristics from ordered-data reduction decisions. Sufficient evidence for addressing these two questions permits more refined analyses of missing data patterns and mechanisms, i.e., user-missing or item non-response, also, as missing-completely, -at, or -not-at random (MCAR, MAR, MNAR; Beale & Little, 1975, Rubin, 1976; c.f., Rao, 1956). More on this later.

First, gleaning the system-missing values from Table 1 above, observe the systematic-reduction of system-missingness at each stratification of the total sample. Less discernible, however, is the *unconditional* system-missing data within units. It is these nonresponse rates that will be of most use in determining who merely failed to be resampled (included once in aggregate file) from whom was unable or refused to respond, that is, what are the completion rates within-unit, given that they started the survey? This is critical to discern selection-borne bias from coverage-borne confounding, that is, missing data by design or by model (see, Little & Rhemtulla,

[35]For the RAND ALP dataset, rank-based panel weighting is implemented to help guard against sampling error. The institute's documentation (overview) of the procedure is provided in Appendix A. On review, it was determined that the weighting employed on select demographics would likely pose negligible issues for the current analyses here, within, and this was independently verified by an institute's statistician. Panel-specific weights have been requested in order to make a more informed determination.

2013) - selection versus confounding due to underrepresentation (adequate coverage These values are depicted in Table 10 below.

For inferences we may wish to make from the sample to the population, the implication is that there is a certain loss of information from conditioning on complete cases of FTP, and this is when conditioning on the joint-probability for all cases (N = 434). Conversely, information across other sample variables is maximized because the conditional marginal probability will be greatest for the most prevalent (systemic) condition (e.g., FTP). Accepting a certain loss, the question becomes whether that information observed is proportional to what we would observe at higher strati, that is, are the relative marginal proportions equivalent? Another question could be, is the information from the conditional and unconditional distributions similar about the first two moments of their distributions?

A simple approach to testing for marginal proportionality would be to simply create binary indicators of responses for all inclusive units (i.e., waves, panels). While the procedure may not be very informative inferentially, it may be quite informative descriptively. That is, it would summarize the marginal probabilities of observing other units in the dataset as a function of responding to the focal FTP unit. These pairwise-tests are presented in Table 3 below. Perhaps unsurprisingly, the crude classifications with fairly large samples are significant. A more precise approach is presented below.

Table 12

Para-data Summary of Unconditional System- and User-missing Data

N Unit-Observations (target constructs)	*N* System-Missing (%)	*N* User-Missing (%)	*N* Complete Cases (%)
4,258 (Future Time Perspective)	27 (< .1)	62 (< .1)	4,169 (> 97.91)
2,922 (Optimism)	17 (< .1)	9 (< .1)	2896 (> 99.1)
2,667 (Subjective Life Expectancy)	0	0	2,667
1,543 (Retirement Plan)	0	0	1,543
1,075 (Sunk Cost, Impulsivity, Choice Deferral)	1, 1, 1 (< .1)	7, 9, 8 (< .1)	1062 (> 99.4)

Table 13

Pairwise Binomial Tests of Unit-Response as Function of FTP Unit-Response

	FTP		
Optimism	Valid	Non-response	φ, *p*-value
Valid	2654	248	
Non-response	1577	1624	.46, $p < .00$
	FTP		
Sunk Costs	Valid	Non-response	φ, *p*-value
Valid	1064	n/a (<5)	
Non-response	3167	1869	.30, $p < .00$
	FTP		
SLE	Valid	Non-response	φ, *p*-value
Valid	1039	25	
Non-response	3192	1847	.28, $p < .00$
	FTP		
Retirement Plan	Valid	Non-response	φ, *p*-value
Valid	1284	35	
Non-response	2497	1837	.32, $p < .00$

Next, item non-responders were factored from the system-missing (unit non-responders) as a function of response to the focal variable, FTP. The relative proportions from variables with the highest percentage of 'system-missingness' were then subjected to Fisher's (1921) contingency test in order to determine if the observed-proportions were statistically comparable, assuming a homogenous population. Stratifications were identified from reductions leading to the base sample, i.e., selection factors (e.g., workers, under 40-years) and were paired for substantive relations to the measured variables. Results in Table 12 indicated non-significant differences and suggested that the conditional observed effects of non-response were statistically comparable, assuming a homogeneous population, also provisional evidence for data missing-at-random, MAR, Little, 1978).[36]

Table 14

Item non-responses probabilities as a function of FTP-unit non-response.

	Retirement	N.R.	Retirement	N.R.	
Workers	803	1727	922	2566	Non-Retirees
(Workers | FTP)	793	1673	904	2453	(Non-Retirees | FTP)
	Fisher's p = .76		Fisher's p = .66		
	Impulsivity	N.R.	Impulsivity	N.R.	
>40	875	2072	1041	3126	All Ages
(>40 | FTP)	837	1939	1039	3101	(All Ages | FTP)
	Fisher's p = .71		Fisher's p = .92		
	SLE	N.R.	SLE	N.R.	
>40	2236	633	3526	641	All Ages
(>40 | FTP)	2228	624	3514	626	(All Ages | FTP)
	Fisher's p = .87		Fisher's p = .76		

Note. N.R. = non-response.

A summary of the user-missing (item non-response) patterns, including the observable variable designated for the structural model, are presented in Table 13. To further examine the nature of the non-response data, a series of Little's MCAR chi-squares (1987) was conducted on combinations of subsamples and measured variable profiles in order to discern plausible patterns approximating missing data mechanisms. Missing values pattern analyses were conducted using the expectation-maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977; Orchard & Woodbury, 1972). Results are presented in Table 14. Only observed variable returned user-missing values, and all were missing > 2%. Little's MCAR (1998) test indicated

[36] Adopting Rubin's (1976) typology, data may be considered MAR if the missing values do not depend on the variable with missing values and can be demonstrated to depend on values of other observables in the dataset, Pr(r | yo, ym) = Pr(r | yo).

non-significance, X^2 (177, N = 384) = 183.80, p = .33, suggesting the missing data mechanism was ignorable (Little & Rubin, 2002; von Hippel, 2004).

Table 15

Summary Statistics for Pattern Analyses of User-Missing Data (Item Non-Response) *in the Base Sample* (N = 434)

Latent Factor (*J* / *K*)	*% Missing*	$\sum_{n=0}^{j(k)} n$ *Univariate Extremes (Low / High)*[1]
Intended Retirement (1)	32.7	3 / 5
SLE (1)	2.3	0 / 0
FTP (6)	0	26 / 0
Optimism (6)	0	61 / 0
Choice Deferral (5)	0	0 / 36
Decisional Impulsivity (4)	0	0 / 13
Sunk Cost (3)	0	0 / 75

Note. J / *K* = # of indicators / latent factor (indicated in parentheses), 1. Univariate extremes defined as > 2 SD + / - $\bar{x}$.

In summary, the missing data section presented information with regards to the total possible information loss as a function of aggregation (Table 1). It was also shown, however, that the relative bias as a result of selection factors may be inflated as a function of the possible sampling space, i.e., conflating item and unit-non response, or, the sampling and responding models for missing data.

Table 16

Summary Results of Little MCAR Chi-square Tests

Focal Variable User-Missingness Mechanism					
Subsample	Variable Composition	[% missing]	*MCAR-*X^2	*Df*	*p-value*
FTP Sys-Mis Rem (*N* = 4,231)	FTP	[0 - .2%]	155.47	148	.32
FTP (Workers) (*N* = 2,516)	FTP	[0 - .2%]	99.63	91	.25
FTP (> 40) (*N* = 2,880)	FTP	[0 - .2%]	102.98	89	.15
FTP (Workers, >40) (*N* = 1,619)	FTP	[0 - .2%]	59.33	59	.46
FTP (Base sample) (*N* = 434)	FTP	n / a	-	-	-
All Variables User-Missingness Mechanism					
Base Sample (*N* = 434)	All Predictors	[0 - 3.7%]	48.77	54	.68
Base Sample (*N* = 434)	All Outcomes	[0 – 33.6%]	14.34	12	.28
Listwise Sample (*N* = 280)	All Predictors	[0 - .4%]	46.03	40	.24
Listwise Sample (*N* = 280)	All Outcomes[1]	-	-	-	-

Note. 1. No outcome variables missing in Listwise sample

In order to account for potential upward bias, Tables 12, 13, and 14 drilled down into the data to separate unit non-response from item non-responses. Evidence indicated that the certain information loss incurred by conditioning on valid responses to FTP was, actually, minimal. In addition, it was shown that non-response to the FTP questionnaire across selection factors indicated evidence of MCAR. Finally, non-responses to constituent variables as a function of item-non response to FTP (factoring out unit non-response) indicated non-significance (provisional evidence for missing-at-random and ignorable mechanisms of the missing data pattern, vis-à-vis, participant item response patterns). Descriptively, it was shown that the amount of missing data on latent indicators was negligible after conditioning on FTP. Conversely, the exogenous measured variables (vectors, $\eta_{(\text{SLE})}$ and $\eta_{(\text{Retirement Plans})}$) exhibited the most missing data. Three viable approaches to this missing data are briefly presented below.

5.4.1.3 Imputation. Cold-deck imputation, as a predecessor to hot-deck imputation, imputes values from prior assessment occasions without unit- or distributional-adjustments to the target sample, i.e., data augmentation. It may be classified as a form of deterministic (non-stochastic) imputation (see, Chapman 1976 for introductory review, also, Rao, 1996). In the current dataset, the single-imputation method (c.f., Rubin, 2008) was attractive primarily because nearly all of the missing values, expectedly, were found for truly exogenous measured variables, i.e., non-latent factor indicators (vectors η). I say expectedly, because a reductionist logic may apply for an exogenous observable variable, whereby, unit non-response approaches system-missing as the number of items, (j) tends toward one. By the converse, the two-phase simple random sampling assumed for the imputation mechanism (item-nonresponse) reduces to simple unit-non response. For the current data, because unit non-response had been previously excluded in the explicit data reduction method, a relaxed assumption was afforded for imputing within-person missing values.

Because the missing value pattern was limited to the dependent column vector variable, $\eta_{(\text{intended retirement})}$, this special case has implications for the expected downward bias in variance estimates and unknown population parameter bias generally introduced with conventional single imputation (and ‘non-proper’ multiple imputation, according to Rubin, 1987) procedures (Anderson, 1979; Bailar & Bailar, 1978; Elofsson, 1973). As Denk and Weber (2011) observe, “Equivalence class matching corresponds to cold-deck donor-based imputation of completely missing variables" (*p*. 10; 2011).

In a partially-naïve, longitudinal real-data scenario, Engels and Diehr (2003) reported findings for the comparative accuracy of within-person imputation methods, relative to population- or class-based imputation. Engels and Diehr (2003) limited their comparative assessment to "single imputations" under "standard analytic methods" (*p. 968*). In a more recent survey of imputation-based strategies for data with nonignorable missing values, Yang, Li, Shoptaw (2008) reported evidence indicating less bias in parameter estimates resulting from within-person, rather than within-panel, info for imputation. With the above-evidence and additional assumptions, it is tenable that within-person values on identical items from different panels is a valid imputation technique. In the present case, respondents with missing value on the intended retirement variable were examined for unit non-response. Then, unit non-responders were cross-checked for valid responses on the initial missing value from another panel. Values from the alternative panel were then examined with matched-cases from the original survey for identifying assumptions permitting statistical inference from candidate imputed values. For the current missing data, this includes a number of distributional assumptions (Bailer & Bailer, 1978). A summary of these tested assumption is reported in Table 15 below. In summary, the findings provide tentative support for the carry-forward method.

Table 17

Summary Descriptives, Non-Parametric, and Parametric Paired-Sample Tests for Retirement Plan Imputation in Base Sample (N = 434)

Summary Descriptives		**Non-Parametric Tests**		**Parametric Tests**			
Ret-Pre $\bar{x}$ (s)	Ret-Foc $\bar{x}$ (s)	Wilcoxin	Marg Homogen	*Pearson-r*	*Paired-t*	*df*	Levene's
65.92 (4.53)	65.61 (4.47)	z = -1.89	z = 1.64	.75**	1.45	238	1.81

Note. Ret-Pre = predated candidate retirement item, Ret-Foc = model retirement variable, Wilcoxin = Wilcoxin signed-rank test, Marg Homogen = marginal homogeneity, Levene's is variance difference test. **p <.01.

5.4.1.4 Direct Maximum Likelihood (DML).[37] Because the degree and variety of data missingness[38] can compromise covariance structural integrity, DML estimation in the context of SEM or, *structured* EM, preserves covariance info through requisite computation of a mean structural model. Pertinent to the current data, the DML method has evidenced more efficient and less biased parameter estimation under weaker distributional assumptions (non-normality), as well as data-missing mechanisms (MAR or MCAR; Enders, 2001; Enders & Bandalos, 2001; Gold, Bentler & Kim, 2003).

5.4.1.5 Listwise-deletion. Provided the negligible amount of missing data on the focal-FTP questionnaire < 1%), as well as evidence for the user-missing data to be MCAR, the Listwise-deletion was deemed a viable approach to missing data handling. In particular, the listwise-deletion approach was viable without conclusive evidence regarding the missing data mechanism owed to sampling and responding. By consequent, the listwise-deletion analyses (i.e., complete-case analysis also afforded a pseudo model-moment sensitivity test to the assumed missing data mechanism. As noted in the IRT analyses, under MCAR assumptions, complete-case analysis reduces power, but yields unbiased parameter estimates.

5.4.2 Measurement Model Specification

All latent variable model analyses were conducted in EQS v6.1 and v6.2, which employs an accessible mathematical and notational model known as the Bentler–Weeks model (1979). Specification of the measurement model followed

---

[37] A form of 'unsupervised' classification algorithm.

[38] Researchers have implicated different attributes of missingness for analytic inferences, including prevalence, severity, mechanism-diagnosable pattern, variety of patterns, among others. Paradigmatic prescriptions for nonnegligible proportions of overall missing data are also varied, e.g., >0% Bentler (2006), >5% Schafer (1999), >10% Bennett (2001), >20% Peng, Harwell,

general procedures from MacCallum and colleagues (1986; MacCallum, Roznowski, & Necowitz, 1992). Specifically, a first-order confirmatory factor analysis (CFA) was conducted in aim of obtaining an independent clusters model, indicating model identification, as well to specify the adequate number of latent factors (Macdonald, 1999).

First, the 5-factor model was estimated and the pattern factor loadings inspected. All items loaded significantly on their conjectured common-factors in the expected direction. Inspection of the distribution of standardized residuals indicated approximate-symmetry centered about zero. In addition, preliminary inspection of the multivariate-Lagrange Multiplier (M-LM) index for potential factor cross-loadings indicated deferred, negligible improved fit for cross-loading estimation (configurally ‘pure’ indicators, McDonald, 1999). However, inspection of the standardized pattern loadings indicated a number of weak-loading items ($\Lambda_{jk} < .32$). Because the goal of the current SEM analyses pertained to testing hypotheses of latent structural coefficients (IRT hypotheses dedicated to psychometric assessment of the focal construct’s measurement), efficient model and parameter estimation was held as premium.

Historically, researchers from the domain sampling tradition have noted that the strategy of data reduction may be utilized in a CFA framework, provided sufficient theoretical specification and adequate sampling of within-domain items (Nunnally, 1978). More recent, many researchers advocate the methodological integration of exploratory, confirmatory, and causal analytic procedures for latent variable modeling (Marsh, Morin, Parker, & Kaur, 2014). Procedurally, because the consideration of item inclusion for subsequent estimation would prohibit log-likelihood model differences as a function of different data, the external misspecification search is conducted prior to model-comparisons (Anderson & Gerbing, 1982).

From a design perspective, the model-estimation within individual-level data (first-order factors) may be facilitated with efficient measurement specification, as well as unidimensional factors. Given the retrospective nature of the study, it could be argued that more latitude be afforded in the specification search. A more substantive reason for conducting the local specification search is owed to simulation and empirical data analyses from MacCallum and colleagues (1999; 2001) indicating the relative import of factor saturation (communality level) for efficient recovery of population-level factors from sample data, compared to sample size or factor-overdeterminacy (j / k) (c.f., Little, Lindenberger, & Nesselroade, 1999). Taken together, item-evaluation proceeded with the goal of retaining indicators of interpretable unidimensional latent factors for structural hypothesis testing (Gerbing & Anderson, 1988; Kano, 2002).

In effort to obtain a clean factor solution for subsequent nested-model comparisons, a heuristic cutoff criteria for conjectured (intended) factor loadings (Λ_{ij}) < .32 determined sequential deletion from low – high on inspection of the standardized loadings pattern (Λ_j . Standardized residuals are also examined (r_{ij} - $\hat{\sigma}_{ij}$ / $s_i s_j$ < 2.0). After each round of eliminations, the measurement model was respecified, estimated, and examined applying the same decisional criteria (Bentler, 2006). A lower-bound of three indicators per latent factor ($3j$ / k) was set to preserve measurement identification (Comrey & Lee, 1992; Tabachnik & Fidell, 2007).[39] Results of the iterative specification search are displayed in Table 16.

[39] A minimum of j = 3 indicators per factor loading was reserved for this procedure.

Estimation results of the base 5-factor measurement model indicated poor global and local fit to the observed sample data, S-Bχ^2= 1780.35 (726, 280), *p* < .00; CFI = .76, AASR = .06, RMSEA = .07 [90% CI = .07, .08]). Inspection of the standardized residuals and pattern factor loadings indicated localized areas of solution strain.

Following respecification, the base 5-factor measurement model was reestimated and results indicated markedly improved fit but, still, only adequate overall fit to the observed sample data, S-Bχ^2= 682.28 (363, 280), *p* < .00; CFI = .90, AASR = .05, RMSEA = .08 [90% CI = .08, .09]). Inspection of the standardized residuals and pattern factor loadings indicated localized areas of solution strain, specifically, with two additional weak-loading items from the fourth factor (Sunk Cost). Following respecification, the base 5-factor measurement model was reestimated and results indicated relatively improved fit to the sample data, indicated primarily by ΔRMSEA, which is a parsimony-favoring (sensitive) index, S-Bχ^2= 615.18 (310, 280), *p* < .00; CFI = .90, AASR = .05, RMSEA = .06 [90% CI = .05, .07]). Inspection of the standardized residuals and pattern factor loadings indicated adequate loadings of retained indicators, with the largest standardized residual contributing to misfit estimated between concatenate items at .27, compared to .60 in the base measurement model. Before proceeding to nested-model comparisons, a consistently-indicated respondent contributing to multivariate nonnormality, on an order > 300%, compared to the second-greatest contributor, was examined in the raw data. There was no detectable response patterns within, or across, response units, likely indicative of a careless, random responder, which would explain the oversight of the consistency-based quality indices. The respondent was eliminated from the dataset.

A series of nested-model comparisons with a profile of measurement fit indices provided complementary info to determine model fit to the sample data (see, Hu & Bentler, 1999; Tanaka, 1993). Specifically, five indices were included as follows: 1) Hierarchical chi-square (X^2), 2) Comparative Fix Index (CFI); Bentler, 1989, 3) Bayesian Information Criterion (BIC); Schwarz, 1978, 4) Root-Mean-Squared-Error of Approximation (RMSEA); Steiger & Ling, 1980, and 5) Average Off-Diagonal Absolute Standardized Residual (AASR), Hu & Bentler (1998). Values are also interpreted following recommendations of Hu and Bentler (1999). Results are presented in Table 17.

Model comparisons began with a single-factor solution and proceeded with the estimation of sequentially added factors. Plausibility of factor-indicator patterns was emphasized (Bagozzi & Yi, 2012). Specifically, a 1-Factor model was specified whereby the metric was set by fixing a factor loading of an item correspondent to the factor with the greatest proportion of indicators in the 5-factor solution (FTP, j = 8).

Table 18

Summary of Restricted Backward-search Measurement Model Respecification in Listwise Sample (N = 279).

Round	Iter	Factor Label	# *J*	$\Lambda_{j,k} < .32$	$<\Lambda_j^2>$	Item Content
1	9	Impulsivity	4	$\Lambda_{5,3} = .27$	.61	When making decisions, I do what seems natural at the moment.
		Sunk Costs	10	$\Lambda_{1,4} = .22$	-	I would waste time worrying about it.
		Sunk Costs	9	$\Lambda_{3,4} = .20$	-	It would take me a long time to adjust myself to it.
		Sunk Costs	8	$\Lambda_{4,4} = .26$	-	I would feel paralyzed.
		Sunk Costs	7	$\Lambda_{5,4} = .17$	-	I would have trouble doing anything at all.
		Sunk Costs	6	$\Lambda_{6,4} = .09$	-	I wouldn't know how to deal.
		Sunk Costs	5	$\Lambda_{7,4} = .31$	.42	I wouldn't have difficulty starting.
		FTP	11	$\Lambda_{10,5} = .13$	-	I have the sense that time is running out.
		FTP	10	$\Lambda_{11,5} = .08$	-	As I get older, I begin to experience time as limited.
		FTP	9	$\Lambda_{12,5} = .01$	-	I feel the importance of time.
		FTP	8	$\Lambda_{7,5} = .05$	.51	I have limited time left to live my life.
2	6	Sunk Costs	4	$\Lambda_{1,4} = .22$	.55	I would take immediate action to correct it.
		Sunk Costs	3	$\Lambda_{3,4} = .20$		I would take action rather than just complaining.
3	5	-	-	-	-	-

Note. Iter = # of Iterations for Optimal Parameter Estimate Convergence. # *J* = remaining items following respecification. $< \Lambda_j^2 >$ = the expected average communality index from remaining indicators.

A 2-Factor model was specified whereby all indicators loaded on either an exogenous (predictor) or endogenous (outcome) latent factor. A 3-Factor model was specified whereby indicators loaded on factors representing each unit of response (wave), that is, one factor for the endogenous factor, and two factors for the predictors determined by the two assessment occasions. A 4-Factor model was specified whereby, again, one factor was designated for each wave of assessment (3), and the wave with the greatest number of items was further specified by item-format (Likert-type and situational judgment). Finally, a 6-Factor model was specified by replicating the factor structure of the conjectured 5-factor solution and, further, by bifurcating the items of the endogenous factor (2 factors for FTP).

Inspection of the nested models indicated retention of the 5-factor measurement model as an adequate solution for representing the observed sample data. It should be noted that, while the chi-square difference test indicated significance for the 6-Factor solution, other indices indicated no change or slightly worse fit, specifically, the parsimony-favoring indices of RMSEA and BIC. In addition, because the ratio-based chi-square difference test is sensitive to sample size, though less-so than standalone chi-squares, it is possible that rejection of the more restricted 5-Factor model is partly due to the sample size (>*N* 200) rather than improved fit to the observed sample data (Hoyle & Panter, 1995). To examine this assumption, an additional 7-Factor model was estimated where the latent predictor with the greatest number of indicators (Optimism, *j* =6) was bifurcated into two, 3-item factors. Results indicated an imperfect solution whereby the chi-square difference actually indicated non-significance, while other goodness of fit indices (residual-based) improved. Still, the information criterion increased, indicating relatively worse fit to the observed data. Stated differently, the standalone chi-square and residual-fit indices improved, while the scaled-difference and information criterion indices indicated non-significance and worse fit, respectively. Because the more parsimonious, 5-Factor solution exhibited comparable residual-based and goodness of fit indices, the 5-Factor model was retained for structural modeling hypothesis testing.

Following latent factor identification, restricted forward specification search proceeded with inspection of the multivariate-Lagrange Multiplier (M-LM) modification index for potential improved fit to the measurement model (Chou & Bentle, 1990; MacCallum, 1986). The M-LM statistics asymptotically approach a chi-square distribution (Bentler & Dijkstra, 1985). The simultaneous procedure was implemented. Statistical alpha is set to .01 for test-wise error control. Also, statistical cues of improved fit were subsequently probed with three substantive deliberations for free estimation rationale: 1) Common-factor loadings of corresponding congeneric indicators (Hooper, 2008), 2) Methodological justification within context of questionnaire responding (Brown & Moore, 2002), and 3) Theoretical soundness of homogeneity for item-content domain (Anderson & Gerbing, 1982). Table 18 summarizes the stepwise-modification decisions[40] and Table 9 displays concordant changes in measurement model fit indices.

[40] It should be noted that there are proponents of modifying one parameter per estimation in spite of the provision of the multivariate-LM index, e.g., Green, Thompson, & Poirer, (1999); Kano & Harada, (2000); Kaplan, (1998); MacCallum, (1986). Following the recommendation of Bentler (2006), the reduced-alpha and pair-wise respecifications may be viewed as a most-conservative compromise to single-sequential modifications.

Table 19

Nested-Measurement Model Comparisons for Listwise Sample (*N* =279)

Model	**S-B X^2**	***Df***	**CFI**	**AASR**	**BIC**	**RMSEA [90% CI]**	**Change Indicators**			
							[1]ΔS-B X^2 / Δ*Df*	**ΔCFI**	**Δ*BIC***	**Δ*RMSEA***
1-Factor	2262.68	324	.40	.11	2567.05	.15 [.14 - .15]				
2-Factor	1883.82	322	.51	.11	2199.16	.13 [.13 - .14]	102.85 / 2*	.11	367.89	.02
3-Factor	1232.46	319	.72	.07	1564.69	.10 [.10 - .11]	125.67 / 3*	.21	634.47	.03
4-Factor	1018.55	315	.78	.06	1373.03	.09 [.08 - .10]	178.01* / 4	.06	191.66	.01
5-Factor	655.39	310	.90	.05	1038.30	.06 [.06 - .07]	226.14* / 5	.12	334.73	.03
6-Factor	621.75	304	.90	.05	1038.44	.07 [.06 - .08]	31.83* / 6	.00	-.14	-.01
7-Factor	613.83	297	.90	.05	1069.41	.06 [.06 - .07]	8.40 / 7	.00	-30.97	.01

Note. * *p* < .00. 1. Scaled S-B chi-squared differences computed. CI = confidence interval.

Table 20

Measurement Model Restricted-Forward Search Respecification Criteria for Listwise Sample (*N* =279)

						Substance		
Round	**M-LM Step**	**θ**	**$X^2_{(Inc)}$**	***p-val***	**Corresponding Indicator Content**	**$\Lambda_{.m}$**	**Meth**	**Thry**
1	1	E6,E5	66.76	<.00	I hardly ever expect things to go my way, I rarely count on good things happening to me.	x	c, r	x
	2	E23,E22	62.11	<.00	Most of my life lies ahead of me, My future seems infinite to me.	x	c	x
2	1	E25,E24	39.57	<.00	I could do anything I want in the future. There is plenty of time in my life to make new plans.	x	c	
	2	E5,E4	32.83	<.00	I hardly ever expect things to go my way, If something can go wrong for me, it will.	x	c, r	x

Note. * $p < .00$. 1. Scaled S-B chi-squared differences computed. CI = confidence interval. $\Lambda_{.m}$ = common-factor of corresponding indicators. $\hat{\theta}$ = parameter estimated for free estimation. $X^2_{(Inc)}$ = univariate increment of modification. Meth = methodological rationale, Thry = item-feature similarity for ordering on theoretical latent factor. x = generic satisfaction of substantive criterion, c = concatenate, r = reverse-score.

Beginning with the base 5-factor model, inspection of the M-LM index indicated free estimation of two indicator error-covariances could improve the model's fit. Application of the decision-criteria permitted the two error-covariance terms to be freely estimated. The modified model was reestimated and, again, the multivariate M-LM modification index was inspected. Modification indices suggested potential improved fit with two additional freely estimated error covariances. Following the same decision criteria, the two error terms were permitted to covary and the model was reestimated (Chou & Bentler, 1996). Model modifications desisted after two rounds of estimation and a total of four covaried error terms in order to reduce over-fitting likelihood (MacCallum, 1986). The modified measurement model was reexamined against a one-factor, four-factor, and six-factor solution, supporting the retention of the 5-factor model (Table 19). The standardize solution for the measurement model in the Listwise sample (N = 279) is displayed in Figure 2. With p = 27 measured variables and k = 72 parameter estimate, the model is over-identified at $(p(p + 1) / 2) – k$ = 378 – 72 with 306 degrees of freedom.

The bivariate correlation matrix with univariate descriptive statistics is displayed in Table 20. A cursory evaluation of the matrix indicated a larger standard deviation for the measured outcome variable, intended retirement, relative to the other variables on Likert-type scales. In addition, perusal of univariate descriptives indicate probable departure from multivariate normality exhibited by values for the third and fourth moments of the distribution.

Table 21

Measurement Model Modification Fit Indices and Configural Comparisons for Listwise Sample (*N* =279)

Model	**S-B X^2**	***Df***	**CFI**	**AASR**	**BIC**	**RMSEA [90% CI]**	**Change Indicators**			
							[1]LR-X^2	**ΔCFI**	**Δ*BIC***	**Δ*RMSEA***
Base	655.39	310	.90	.05	1038.30	.06 [.06 - .07]				
2-E Cov	548.52	308	.93	.05	942.42	.05 [.05 - .06]	102.85 / 2*	.03	95.88	.01
4-E Cov	482.83	306	.95	.05	888.26	.04 [.04 - .05]	79.40 / 2*	.02	54.16	.01

Model	**S-B X^2**	***Df***	**CFI**	**AASR**	**BIC**	**RMSEA [90% CI]**	**Change Indicators**			
							[2]ΔS-B X^2/ ΔDf	**ΔCFI**	**ΔBIC**	**ΔRMSEA**
4-E 1 Fac	2038.97	320	.46	.10	2365.57	.16 [.13 - .15]	[3]1156.41 / 14*	.49	95.88	.12
4-E 4 Fac	858.28	311	.83	.06	1235.56	.08 [.07 - .09]	217.11 / 14*	.49	95.88	.12
4-E 6 Fac	469.30	300	.95	.08	908.52	.05 [.04 - .05]	12.87 / 6	.04	54.16	.04

Note. * $p < .00$. 1. LM - Modification Index Computed on Likelihood Ratio Chi-squared Differences. 2. Scaled-Satorra Bentler Chi-Squared Differences Computed. 3. Change Indicators Computed in Reference to Modified Base 5-Factor Model. CI = confidence interval. 2-E = Two errors covaried, 4-E = Four errors covaried. Fac = Factor.

Data exhibited mixed-skewness, while the majority of greatest departures from univariate normality (> |1|) were negatively skewed. More relevant to the covariance structure analyses, the majority of variables indicated slight platykurtosis, while the two greatest departures from univariate normality were leptykurtotic (> 1). Further, the normalized estimate of Mardia's Kappa coefficient for multivariate normality was 22.21.

The robust method to maximum-likelihood (R-ML) is used for model estimation for its robustness to certain distributional assumptions (Mardia, 1970), as well as ignorability of the missing data mechanism. This method tends to yield less biased fit estimates than asymptotic distribution-free (ADF) estimators when sample sizes are relatively small ($N < 2000$) (Gold, Bentler, & Kim, 2002). The Yuan-Bentler residual-based chi-square (YB-X^2) with correctly scaled X^2-difference tests and adjusted standard error estimates for individual parameters will be interpreted, as well as the Yuan-Bentler residual-based F-test, due to the former statistic's conservatism in small samples (Yuan & Bentler, 1998; Bentler & Yuan, 1999).

Table 22.

Summary Descriptive Statistics of All Observable Variables for Listwise Sample

Variable	*M*	*SD*	*Range*	*Skew*	*Kurtosis*
Optimism 1	4.57	1.25	5	-1.01	.53
Optimism 2	4.28	1.17	5	-.69	.14
Optimism 3	4.81	1.18	5	-1.13	1.14
Optimism 4	4.56	1.49	5	-.70	-.63
Optimism 5	4.84	1.23	5	-1.00	.40
Optimism 6	4.84	1.34	5	-1.08	.44
Choice Deferral 1	2.04	.97	4	.78	.00
Choice Deferral 2	1.85	.97	4	.98	.23
Choice Deferral 3	2.32	1.08	4	.49	-.60
Choice Deferral 4	2.08	1.00	4	.67	-.31
Choice Deferral 5	1.89	.98	4	1.03	.36
Decisional Impulsivity 1	2.06	.95	4	.53	-.49
Decisional Impulsivity 2	2.24	1.01	4	.37	-.62
Decisional Impulsivity 3	2.75	1.04	4	-.14	-.74
Decisional Impulsivity 4	2.16	.92	4	.46	-.36
Sunk Costs 1	.09	.01	4	.91	.35
Sunk Costs 2	.61	.08	4	.22	-.57
Sunk Costs 3	.56	.17	4	.49	-.56
Future Time Perspective 1	4.92	1.53	6	-.31	-.69
Future Time Perspective 2	4.57	1.6	6	-.16	-.83
Future Time Perspective 3	5.1	1.53	6	-.55	-.38
Future Time Perspective 4	3.61	1.51	6	.25	-.48
Future Time Perspective 5	3.73	1.75	6	.28	-.82
Future Time Perspective 6	4.27	1.63	6	.01	-.94
Future Time Perspective 7	4.3	1.62	6	.04	-.93
Future Time Perspective 8	4.87	1.67	6	-.43	-.89
Intended Retirement	65.82	4.51	35	.32	1.82

Table 23

Bivariate Correlation Matrix and Summary Descriptive Statistics for the Listwise Sample (*N* = 279)

	Op1	Op2	Op3	Op4	Op5	Op6	CD1	CD2	CD3	CD4	CD5	DI1	DI2	DI3	DI4	SC1	SC2	SC3
Op1	1																	
Op2	.66	1																
Op3	.64	.66	1															
Op4	.35	.43	.46	1														
Op5	.46	.54	.58	.61	1													
Op6	.45	.46	.55	.46	.72	1												
CD1	-.15	-.15	-.11	-.13	-.24	-.18	1											
CD2	-.14	-.16	-.07	-.12	-.19	-.17	.64	1										
CD3	-.07	-.05	-.02	-.10	-.08	-.06	.49	.53	1									
CD4	-.05	-.01	.02	-.03	-.06	-.10	.52	.48	.66	1								
CD5	-.11	-.15	-.05	-.06	-.12	-.11	.46	.49	.55	.54	1							
DI1	-.02	-.02	.03	.05	-.07	-.05	.17	.23	.17	.34	.19	1						
DI2	.02	.04	.08	.07	.01	.02	.22	.25	.18	.36	.17	.77	1					
DI3	.15	.04	.09	.13	.06	.06	-.05	-.06	-.01	.09	-.01	.50	.55	1				
DI4	-.13	-.14	-.02	.03	-.10	-.05	.26	.26	.21	.28	.31	.50	.61	.43	1			
SC1	-.17	-.17	-.21	-.03	-.09	-.14	.25	.24	.20	.20	.34	.02	.00	-.14	.10	1		
SC2	-.18	-.18	-.20	-.04	-.09	-.09	.18	.26	.22	.20	.29	.01	-.02	-.12	.07	.67	1	
SC3	-.18	-.20	-.18	-.12	-.04	-.02	.17	.17	.19	.14	.23	-.03	-.08	-.16	.07	.51	.56	1

Table 21 (continued)

F1	.29	.33	.33	.29	.22	.22	-.06	-.08	-.09	-.01	-.10	.05	.11	.19	.02	-.12	-.08	-.12	1								
F2	.24	.32	.25	.22	.21	.26	-.09	-.10	-.12	-.10	-.16	.03	.04	.13	-.08	-.04	.00	-.06	.66	1							
F3	.29	.35	.31	.32	.27	.26	-.05	-.06	-.06	-.02	-.10	.07	.10	.14	-.02	-.09	-.06	-.15	.75	.74	1						
F4	.25	.31	.19	.14	.12	.13	-.19	-.19	-.19	-.11	-.16	-.01	-.01	.02	-.17	-.10	-.04	-.04	.41	.42	.49	1					
F5	.23	.32	.19	.25	.14	.15	-.14	-.12	-.14	-.06	-.11	.02	.11	.09	.01	-.15	-.15	-.16	.47	.40	.52	.65	1				
F6	.23	.28	.15	.16	.11	.11	-.02	-.12	-.09	.04	-.19	.12	.11	.16	.01	-.11	-.08	-.16	.46	.48	.56	.45	.48	1			
F7	.17	.23	.19	.15	.15	.20	-.08	-.09	-.06	.00	-.15	-.02	.03	.07	-.03	-.12	-.06	-.10	.48	.53	.52	.43	.46	.62	1		
F8	.24	.25	.15	.26	.21	.20	-.15	-.16	-.15	-.08	-.22	-.10	-.01	.04	-.11	-.04	-.02	-.05	.56	.43	.52	.41	.43	.40	.41	1	
RP	-.02	.00	-.01	-.01	-.02	-.02	.03	.02	-.03	.02	.00	.00	-.01	-.04	-.01	-.02	.11	.12	-.01	.06	.05	-.03	-.04	-.06	-.04	-.08	1

Note. Opt = Optimism, Def = Choice Deferral, Imp = Impulsivity, SC = Sunk Costs, RtPln = Retirement Plan, Skw = Skew, Krt = Kurtosis, Ran = Range.

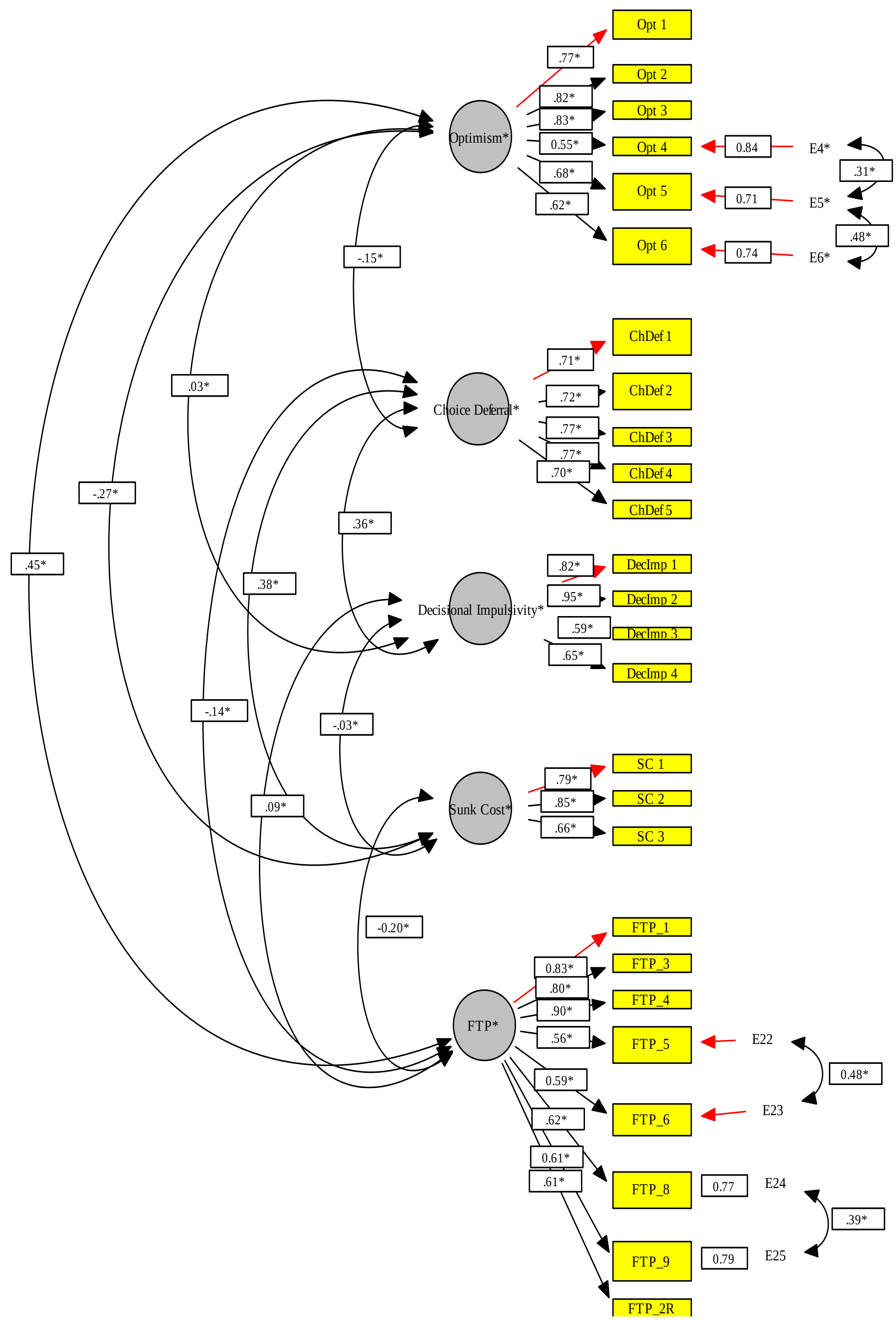


Figure 16. Standardized solution for measurement model for listwise sample (N = 279)

5.5 Retro-Obs Results (SEM)

5.5.1 Structural Model Estimation for Main-Effect Hypothesis Testing

In order to test for main effects hypotheses, a single-group recursive structural model was estimated. In the model, latent factor FTP and measured variable, intended retirement, were simultaneously regressed on the four exogenous latent predictors. Structural coefficients (latent paths) were inspected between exogenous latent variables and endogenous latent and observed variables, respectively. Unstandardized structural coefficients with robust standardized errors in parantheses are reported. Convergence and estimation problems were checked, as well as standardized residual distributions, prior to statistical interpretation of parameter estimates. LM and Wald tests for potential structural model modification was included in the analysis. The LM-test specified the factor – factor structural relations from the gamma submatrix. Global model fit indices are also reported.

Initial inspection of the estimated structural model indicated a fairly well-fitting model to the observed sample data with robust estimates of SB-$X^2_{(307)}$ = 483.49, AASR = .05, RMSEA = .05, CFI = .95, BIC = 883.300. The distribution of standardized residuals also appeared approximately symmetric centered about zero. For the DML-estimated sample (N = 434) with missing values on measured variable intended retirement, the auto-generated generalized least squares-based (GLS) tests of homogeneity of means and covariances for the missing-values pattern indicated non-significance, $GLS_{(\bar{x})}$ X^2 (26, 434) = 36.86, p = .08, $GLS_{(\Sigma i,j)}$ X^2 (351, 434) = 391.65, p = .07. The finding provides support for the missing-value cases and complete cases being sampled from a similarly homoegenous population, also, it permits inferences of parameter estimates obtained from the DML estimation method in a single structural model. That is, there is evidence that any non-response bias in structural parameter estimates is negligible.

R-H1a**:** Optimism will positively relate to FTP.
R-H1b**:** Optimism will positively relate to intended retirement.

Inspection of the latent path coefficient between latent variables Optimism and FTP indicated a significant positive parameter estimate $\gamma_{51} = .62$ (.13), $t(351) = 4.82$, $p < .00$, providing support for hypothesis 1a. Inspection of the structural path coefficient between latent variable Optimism and measured variable Intended Retirement was in the hypothesized direction, but indicated non-significance and providing no support for hypothesis 1b, $\gamma = .05$ (.30), $t(351) = .16$, $p > .05$.

R-H2a**:** Choice deferral will negatively relate to FTP.
R-H2b**:** Choice deferral will negatively relate to intended retirement.

Inspection of the structural path coefficient between latent variables Choice Deferral and FTP indicated a significant negative parameter estimate $\gamma_{52} = -.28$ (.14), $t(351) = -1.96$, $p < .00$, providing support for hypothesis 2a. The value was found to be non-significant in the base-DML sample, $\gamma_{52} = -.13$ (.09), $t(351) = -1.32$, $p > .05$. Inspection of the structural path coefficient between latent variable Choice Deferral and measured variable Intended Retirement was in the hypothesized direction, but indicated non-significance and provided no support for hypothesis 2b, $\gamma = -.19$ (.63), $t(351) = -.29$, $p > .05$.

R-H3a**:** Decisional impulsivity will positively relate to FTP.

R-H3b**:** Decisional impulsivity will positively relate to intended retirement.

Inspection of the structural path coefficient between latent variables Decisional Impulsivity and FTP indicated a non-significant parameter estimate trending in the hypothesized direction, but providing no support for hypothesis 3a, γ_{53} = .18 (.10), $t(351) = 1.89$, $p = .07$.[41] Inspection of the structural path coefficient between latent variable Decisional Impulsivity and measured variable Intended Retirement indicated no relation, γ = .00 (.38), $t(351) = .00$, $p > .05$), and failing to provide support for hypothesis 3b.

R-H4a**:** Sunk costs will negatively relate to FTP.

R-H4b**:** Sunk costs will negatively relate to intended retirement.

Inspection of the structural path coefficient between latent variables Sunk Cost and FTP indicated a weak relationship, but in counter-direction to the hypothesis, γ_{54} = .08 (.11), $t(351) = .67$, $p > .05$, providing no support for hypothesis 4a. Inspection of the structural path coefficient between latent variable Sunk Cost and measured variable Intended Retirement was also non-significant, further, counter to the hypothesized direction and providing no support for hypothesis 4b, γ = .49 (.40), $t(351) = 1.25$, $p > .05$. The obtained parameter trended toward significance in the base-DML sample, γ = .74 (.41), $t(351) = 1.80$, $p = .07$.

---

[41] Prior to subsequent hypothesis testing in larger samples, a least-squares variant of the computed reweighted matrix based on heterogeneous kurtosis distribution assumptions (non-normal theory) yielded an estimator indicating significance of Decisional Impulsivity as a predictor of FTP at the conventional $\alpha = .05$ level, γ = .19 (.08) (Kano, Berkane, & Bentler, 1990). All other structural relations obtained were consistent with the ML estimation method in terms of statistical significance.

Inspection of the standardized solution indicated approximately 22.3% of the observed variance in FTP was accounted for by the hypothesized latent predictors. In contrast, very little of the variance in intended retirement was accounted for by the hypothesized predictors (approximately 1%). It is possible that the misprediction is owed to epiphenomenal Brunswickian asymmetry so that the vector may be poorly predicted by multiple-indicator composites. The possibility is suggested by the large variance estimate for the intended retirement variable. In addition, the significant unexplained variance in the latent outcome variable may suggest room for differential prediction. Before presenting tests of the formal moderator hypotheses, a brief note on the approach to latent interactions is warranted.

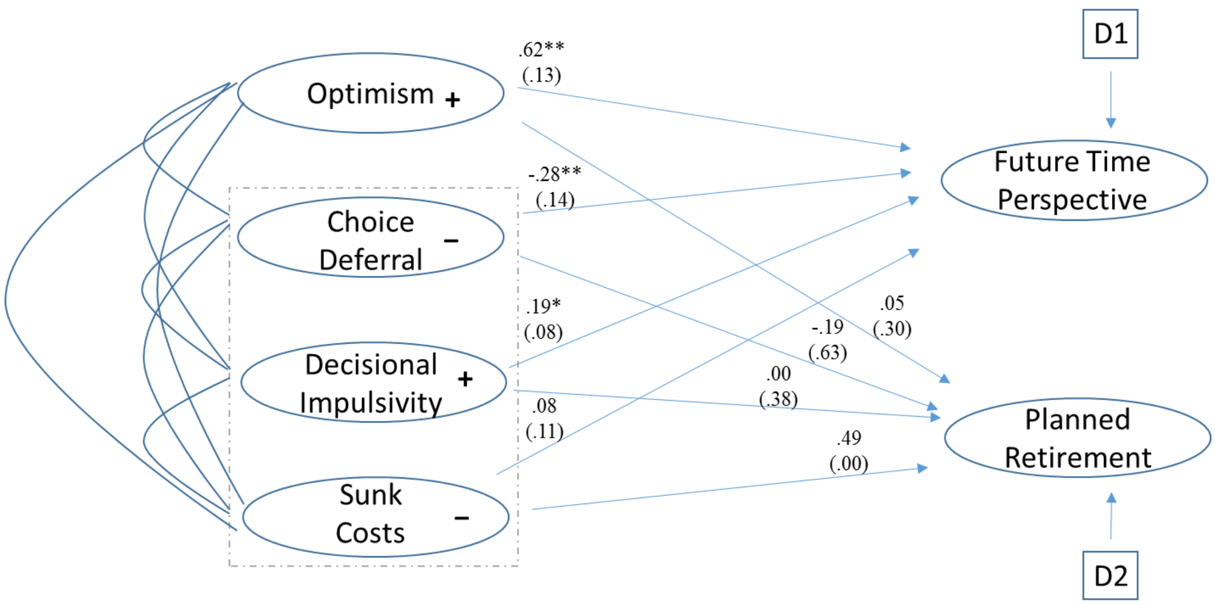


Figure 17. Summary of Obtained Main-Effect Structural Paramter Estimates. +p < .10, *p < .05, **p < .01.

5.5.2 Approach to Latent Interactions

Many of the interaction estimation approaches applieded in OLS frameworks have been extended to latent variable modeling (Jaccard & Wan, 1995, see, for an overview, Steinmetz, Davidov, & Schmidt 2011). Approaches to modeling latent interactions may be broadly categorized into multiple-group (Bagozzi & Yi, 1989) and single-group techniques[42] (see, Hoyle, 2014; c.f., Ping, 1995; 2006[43]). Typically, the multiple-groups approach lends itself to data with a categorical differentiator so that groups may be formed without inherent information loss from artificial categorization of continuous variables, e.g., median-splits. Given the premium for model and parameter estimation efficiency, however, the multiple-groups approach was adopted in order to utilize full structural model information in tests of moderation (Bagozzi & Yi, 1991). The multiple-groups approach to moderation testing assesses invariance of the hypothesized structural coefficient, specifically, by imposing an equality constraint on the latent path coefficient and conducting a likelihood-ratio test of improved fit as a function or releasing the constraint.

[42] The single-group approach comprises many estimation strategies, with little consensus, including latent product indicators, distribution-analytic, and Bayesian-based methods.

[43] Robert Ping technically classifies multiple-group approaches to latent interaction estimation under the label '*indirect'*, distinguished from '*direct*' estimation techniques that necessarily yield structural coefficient estimates without additional "convenience" (measurement) variables in the structural model. In other words, the multiple-groups approach is designated as model *specification* under the '*indirect*' label for estimation. It should be noted, however, that the single- and multiple-*indicator specifications* are fully crossed with the *direct* and *indirect* superordinate labels of estimation. An elaborated taxonomical exposé is beyond the purview of the current dissertation, but, to the extent that estimation may accord with analysis, the current author concurs sampling (multi-group) to be more consonant with measurement (see, Heckman [1979]).

5.5.2.1 Structural Model Estimation for Interaction-Effect Hypothesis Testing. Low and high subjective life expectancy (SLE) groups were constructed based on a median split (50%) on the proposed moderator. Raw data from the low-SLE (*N* = 149) and high-SLE (*N* = 131) groups were used to obtain estimates for the multiple-group SEM analysis. Equality constraints were imposed on factor loadings and structural coefficients, only, permitting latent factor variance and covariance, as well as error-covariance estimates to vary across groups. Summary descriptive statistics for each group on all measured variables are displayed in Table 21 below.

Initial inspection of the multigroup structural model predicting FTP indicated a fairly well-fitting model to the observed sample data with robust estimates of SB-$X^2_{(635)}$ = 860.55, AASR = .06, RMSEA = .05, CFI = .93. Standardized residuals also appeared approximately symmetric and centered about zero, and output indicated that all equality constraints were correctly imposed and parameter estimates obtained without particular challenges to optimization.

R-H6a**:** SLE will interact with Optimism in predicting FTP, such that higher SLE will increase the positive effect.

After imposing an equality constraint on the latent path coefficient between Optimism and FTP, the structural model was re-estimated. Inspection of the global-fit index indicated a non-significant decrement to overall model fit, ΔSB-$X^2_{(1)}$ = 0.26, *p* = .71, ΔAASR = .00, ΔRMSEA = .00, ΔCFI = .00. The finding provided no support for hypothesis 6a.

***R-H7a*:** SLE will interact with Choice Deferral in predicting FTP, such that higher SLE will buffer the negative effect.

After imposing an equality constraint on the latent path coefficient between Choice deferral and FTP, the structural model was re-estimated. Inspection of the global-fit index indicated a non-significant decrement to overall model fit, ΔSB-$X^2_{(1)}$ = 0.20, p = .82, ΔAASR = .00, ΔRMSEA = .00, ΔCFI = .00. The finding provided no support for hypothesis 7a.

***R-8a*:** SLE will interact with Decisional Impulsivity in predicting FTP, such that higher SLE will increase the positive effect.

After imposing an equality constraint on the latent path coefficient between Choice defferal and FTP, the structural model was re-estimated. Inspection of the global-fit index indicated a non-significant decrement to overall model fit, ΔSB-$X^2_{(1)}$ = 0.41, p = .55, ΔAASR = .00, ΔRMSEA = .00, ΔCFI = .00. The finding provided no support for hypothesis 8a.

***R-H9a*:** SLE will interact with Sunk Costs in predicting FTP, such that higher SLE will buffer the negative effect.

After imposing an equality constraint on the latent path coefficient between Sunk Costs and FTP, the structural model was re-estimated. Inspection of the global-fit index indicated a non-significant decrement to overall model fit, ΔSB-$X^2_{(1)}$ = 0.42, p =.53, ΔAASR = .00, ΔRMSEA = .00, ΔCFI = .00. The finding provided no support for hypothesis 9a.

R-H6b**:** SLE will interact with Optimism in predicting intended retirement, such that higher SLE will buffer the positive effect.

After imposing an equality constraint on the latent path coefficient between Optimism and intended retirement, the structural model was re-estimated. Inspection of the global-fit index indicated a non-significant decrement to overall model fit, ΔSB-$X^2_{(1)}$ = 0.18, p =.91, ΔAASR = .00, ΔRMSEA = .00, ΔCFI = .00. The finding provided no support for hypothesis 6b.

R-H7b**:** SLE will interact with Choice Deferral in predicting intended retirement, such that higher SLE will increase the negative effect.

After imposing an equality constraint on the latent path coefficient between Choice Deferral and intended retirement, the structural model was re-estimated. Inspection of the global-fit index indicated a non-significant decrement to overall model fit, ΔSB-$X^2_{(1)}$ = 0.42, p =.53, ΔAASR = .00, ΔRMSEA = .00, ΔCFI = .00. The finding provided no support for hypothesis 7b.

R-H8b**:** SLE will interact with Decisional Impulsivity in predicting intended retirement, such that higher SLE will increase the negative effect.

After imposing an equality constraint on the latent path coefficient between Decisional Impulsivity and FTP, the structural model was re-estimated. Inspection of the global-fit index indicated a non-significant decrement to overall model fit, ΔSB-$X^2_{(1)}$ = 0.21, p =.83, ΔAASR = .00, ΔRMSEA = .00, ΔCFI = .00. The finding provided no support for hypothesis 8b.

***R-H9b*:** SLE will interact with Sunk Costs in predicting intended retirement, such that higher SLE will buffer the positive effect.

After imposing an equality constraint on the latent path coefficient between Sunk Costs and FTP, the structural model was re-estimated. Inspection of the global-fit index indicated a non-significant decrement to overall model fit, ΔSB-$X^2_{(1)}$ = 0.08, p =.97, ΔAASR = .00, ΔRMSEA = .00, ΔCFI = .00. The finding provided no support for hypothesis 9b.

Table 24

Summary of Obtained Structural Parameter Estimates Across Samples, Measurement Model Specification and Structural Model Estimation Methods

Sample	Listwise (*N* =279)		Base (*N =434*)			
Measurement Model Specification	Deletion	Parcel	Deletion		Parcel	
Structural Model Estimation	ML, HKLS	ML	DML	Cold-ML	DML	Cold-ML
Main-Effect Hypotheses						
1a) Optimism → FTP	γ = .62 (.13)**	γ = .43 (.10)**	γ = .49 (.09)**	γ = .50 (.07)**	γ = .27 (.06)**	γ = .41 (.07)**
2a) Choice Deferral → FTP	γ = -.28 (.14)*	γ = -.69 (.30)**	γ = -.13 (.10)	γ = -.13 (.10)	γ = -.38 (.23)$^+$	γ = -.52 (.22)*
3a) Impulsivity → FTP	γ = .18 (.10)$^+$	γ = .48 (.21)**	γ = .06 (.08)	γ = .05 (.08)	γ = .26 (.15)$^+$	γ = .28 (.16)$^+$
4a) Sunk Cost → FTP	γ = .08 (.11)	γ = -.02(.20)	γ = -.08 (.09)	γ = -.07 (.09)	γ = -.29 (.14)$^+$	γ = -.14 (.15)
1b) Optimism → Ret Plan	γ = .05 (.30)	γ = -.04 (.15)	γ = .21 (.30)	γ = .13 (.37)	γ = .07 (.14)	γ = -.04 (.20)
2b) Choice Deferral→ Ret Plan	γ = -.19 (.63)	γ = .01 (.58)	γ = -.27 (.58)	γ = -.23 (.57)	γ = -.08 (.64)	γ = -.17 (.56)
3b) Imp→ Ret Plan	γ = .00 (.38)	γ = .00 (.39)	γ = .04 (.40)	γ = -.07 (.40)	γ = .05 (.41)	γ = -.04 (.42)
4b) SC→ Ret Plan	γ = .49 (.40)	γ = .11 (.33)	γ = .74 (.41)	γ = .50 (.40)	γ = .28 (.36)	γ = .10 (.33)
Interaction-Effect Hypotheses						
1a) Optimism → FTP | SLE	$\Delta SB\text{-}X^2$ = .26	$\Delta SB\text{-}X^2$ = .20	$\Delta SB\text{-}X^2$ = .39	-	$\Delta SB\text{-}X^2$ = 4.89*	-
2a) Choice Deferral → FTP | SLE	$\Delta SB\text{-}X^2$ = .20	$\Delta SB\text{-}X^2$ = 1.40	$\Delta SB\text{-}X^2$ = 1.13	-	$\Delta SB\text{-}X^2$ = 3.64*	-
3a) Impulsivity → FTP | SLE	$\Delta SB\text{-}X^2$ = .41	$\Delta SB\text{-}X^2$ = 1.43	$\Delta SB\text{-}X^2$ = .74	-	$\Delta SB\text{-}X^2$ = 4.94*	-
4a) Sunk Cost → FTP | SLE	$\Delta SB\text{-}X^2$ = .42	$\Delta SB\text{-}X^2$ = 1.22	$\Delta SB\text{-}X^2$ = .39	-	$\Delta SB\text{-}X^2$ = .56	-
1b) Optimism → Ret Plan | SLE	$\Delta SB\text{-}X^2$ = .18	$\Delta SB\text{-}X^2$ = .45	$\Delta SB\text{-}X^2$ = .38	-	$\Delta SB\text{-}X^2$ = .28	-
2b) Choice Defer → Ret Plan | SLE	$\Delta SB\text{-}X^2$ = .42	$\Delta SB\text{-}X^2$ = .78	$\Delta SB\text{-}X^2$ = .29	-	$\Delta SB\text{-}X^2$ = .53	-
3b) Impulsivity → Ret Plan | SLE	$\Delta SB\text{-}X^2$ = .21	$\Delta SB\text{-}X^2$ = n/a	$\Delta SB\text{-}X^2$ = 1.13	-	$\Delta SB\text{-}X^2$ = .37	-
4b) Sunk Cost → Planned Ret |SLE	$\Delta SB\text{-}X^2$ = .08	$\Delta SB\text{-}X^2$ = .08	$\Delta SB\text{-}X^2$ = 2.4$^+$	-	$\Delta SB\text{-}X^2$ = .43	-

Note. ^+p < .10 *p < .05, **p < .01. Unstandardized coefficients with ML-robust standard errors in parentheses reported. All $\Delta SB\text{-}X^2$ values are scaled and estimated with 1 degree of freedom.

5.5.3 Sample Contrasts

In order to ascertain what impact, if any, the selection factors for the focal sample may have had on model parameter estimates, a separate set of analyses was conducted on samples grouped by these selection factors.[44] To optimize comparability of statistical tests conducted on the focal sample, all other sample characteristics were approximately equated. Specifically, two contrast-samples were formed on selection factors, 'chronological age and 'job status'. Contrast analyses based on 'chronological age' comprised a sample of workers over age-45 years (N = 399) and workers under age-46 (N = 120). Contrast analyses based on 'job status' comprised workers of all ages (N = 519) and non-workers of all ages (N = 162).[45]

Measurement model specification included both restricted-backward specification (reduction) and parceling procedures (see, Appendix D for parceling procedure results). Structural model estimation was limited to Listwise and DML treatments of missing data, which was considered an efficient approach to comprehensive analyses for the contrast samples, specifically, with the weakest assumptions regarding models of missing data mechanisms. Results are presented for each contrast analysis below. For space considerations of the main-body document, only structural analyses are presented. All measurement model tables, figures, and descriptive interpretations of concomitant statistics tests are presented in Appendix C.

---

[44] Common terminology for such analyses include, *inter alia,* subgrouping, blocking, or stratifying.

[45] Note, strict cutoffs varied slightly from original samples. For example, in the 'age-contrast' sample, the original 40-years cutoff was relaxed to 45-years in order to ensure adequate sample sizes for obtaining more reliable measurement and structural parameter estimates in the latent-analytic framework. Similarly, for the 'job status-contrast' sample, the non-workers job status is a combination of all job statuses, excluding retirees. Exclusion of retirees aimed to optimize comparability of the criterion, 'planned retirement'. In addition, the age-criteria was also relaxed for this subsample due to the primary composition of workers and retirees in the panel study (>80%, combined).

5.5.3.1 Age contrast – structural model. Results are presented for the listwise sample. Discrepancies of significance and non-significance in terms of obtained findings for the listwise sample are reported and indicated as those obtained from the DML-estimated base sample.

Initial inspection of the estimated structural model indicated a fairly well-fitting model to the observed data in the younger worker sample with robust estimates of SB-X^2(238, N =178) = 339.91, $p < .05$, AASR = .05, RMSEA = .05, CFI = .95. The same structural model was slightly worse-fitting in the older worker sample, but overall, was still a fairly well-fitting to the observed sample data, SB-X^2(238, N =139) = 338.35, $p < .05$, AASR = .06, RMSEA = .06, CFI = .91. Again, for the DML-estimated sample (N = 519), inspection of the generalized least squares-based (GLS) tests of homogeneity of means and covariances for the missing-value pattern indicated non-significance in both the younger worker, $GLS_{(\bar{x})}$ X^2 (23, 120) = 25.85, $p = .31$, $GLS_{(\Sigma i,j)}$ X^2 (276, 120) = 303.65, $p = .12$, and older worker samples, $GLS_{(\bar{x})}$ X^2 (40, 398) = 46.67, $p = .36$, $GLS_{(\Sigma i,j)}$ X^2 (276, 398) = 304.13, $p = .12$.

R-H1a**:** Optimism will positively relate to FTP.
R-H1b**:** Optimism will positively relate to intended retirement.

Inspection of the latent path coefficient between latent variables Optimism and FTP in the younger worker sample indicated a significant positive parameter estimate γ_{51} = .61 (.13), t(238) = 4.77, $p < .00$. Inspection of the corresponding latent path in the older worker sample also indicated a significant parameter estimate in the hypothesized direction, γ_{51} = .42 (.21), t(238) = 1.98, $p < . 03$.

Inspection of the structural path coefficient between latent variable Optimism and measured variable Intended Retirement in the younger worker sample indicated a negligible relationship, $\gamma_{ret,1}$ = .01 (.49), $t(238) = .02$, $p > .05$. Inspection of the corresponding path coefficient in the older worker sample indicated a comparatively stronger positive coefficient, but non-significant, $\gamma_{ret,1}$ = .27 (.34), $t(238) = .78$, $p >$ .05.

R-H2a**:** Choice deferral will negatively relate to FTP.

R-H2b**:** Choice deferral will negatively relate to intended retirement.

Inspection of the latent path coefficient between latent variables Choice Deferral and FTP in the younger worker sample indicated a non-significant negative parameter estimate, γ_{52} = -.12 (.14), $t(238) = -.83$, $p > .05$. Inspection of the corresponding latent path in the older worker sample indicated a similar, non-significant negative parameter estimate, γ_{52} = -.06 (.20), $t(238) = -.31$, $p > .05$. The parameter estimated with DML in the older worker sample trended toward significance in the hypothesized direction, γ_{52} = -.14 (.10), $t(298) = -1.39$, $p = .08$.

Inspection of the structural path coefficient between latent variable Choice Deferral and measured variable Intended Retirement in the younger worker sample indicated a significant negative parameter estimate, $\gamma_{ret,2}$ = -1.25 (.53), $t(238) = -2.36$, $p < .01$. Inspection of the corresponding path coefficient in the older worker sample indicated a non-significant parameter estimate but, interestingly, in the opposite direction and contrary to the hypothesis, γ_{52} = 1.10 (.80), $t(238) = 1.37$, $p = .09$.

R-H3a**:** Decisional impulsivity will positively relate to FTP.

R-H3b**:** Decisional impulsivity will positively relate to intended retirement.

Inspection of the latent path coefficient between latent variables Decisional Impulsivity and FTP in the younger worker sample indicated a non-significant positive parameter estimate γ_{53} = .11 (.12), $t(238) = .87$, $p > .05$. Inspection of the corresponding latent path in the older worker sample indicated a similar non-significant positive parameter estimate, γ_{53} = .03 (.13), $t(238) = .19$, $p > .05$.

Inspection of the structural path coefficient between latent variable Decisional Impulsivity and measured variable Intended Retirement in the younger worker sample indicated a positive, non-significant parameter estimate, $\gamma_{ret,3}$ = .59 (.48), $t(238) = 1.23$, $p = .11$. Inspection of the corresponding path coefficient in the older worker sample also indicated a non-significant path coefficient, but in the opposite direction than hypothesized, $\gamma_{ret,3}$ = -.63 (.42), $t(238) = -1.49$, $p = .11$.

***R-H4a*:** Sunk costs will negatively relate to FTP.
***R-H4b*:** Sunk costs will negatively relate to intended retirement.

Inspection of the latent path coefficient between latent variables Sunk Costs and FTP in the younger worker sample indicated a negligible parameter estimate γ_{54} = .00 (.11), $t(238) = .01$, $p > .05$. Inspection of the corresponding latent path in the older worker sample indicated a non-significant negative parameter estimate, γ_{54} = -.15 (.18), $t(238) = -.89$, $p > .05$.

Inspection of the structural path coefficient between latent variable Sunk Costs and measured variable Intended Retirement in the younger worker sample indicated a significant positive parameter estimate, $\gamma_{ret,4}$ = 1.09 (.46), $t(238) = 2.36$, $p < .05$. Inspection of the corresponding path coefficient in the older worker sample indicated a non-significant positive parameter estimate, $\gamma_{ret,4}$ = .13 (.72), $t(238) = .19$, $p > .05$. Interestingly, the reversed-pattern in terms of significance, but commensurate directional findings, was indicated in the DML-estimates. That is, the parameter estimate was significant in the older work sample $\gamma_{ret,4}$ = .77 (.41), $t(276) = 1.96$, $p < .05$ and non-significant in the younger worker sample, $\gamma_{ret,4}$ = .47 (.86), $t(276) = .54$, $p < .05$.

5.5.3.2 Job status - contrast – structural model. Reporting of results follows the same convention as stated in the Age-contrast structural model above.

***R-H1a*:** Optimism will positively relate to FTP.
***R-H1b*:** Optimism will positively relate to intended retirement.

Inspection of the latent path coefficient between latent variables Optimism and FTP in the worker sample indicated a significant positive parameter estimate γ_{51} = .50 (.10), $t(309) = 5.15$, $p < .00$. Inspection of the corresponding latent path in the non-retiree sample also indicated a significant parameter estimate in the hypothesized direction, γ_{51} = .43 (.18), $t(309) = 2.44$, $p < .00$.

Inspection of the structural path coefficient between latent variable Optimism and measured variable Intended Retirement in the worker sample indicated a positive non-significant parameter estimate, $\gamma_{ret,1}$ = .07 (.36), $t(309) = .20$, $p > .05$. Inspection of the corresponding path coefficient in the non-retiree sample indicated a negative non-significant coefficient, $\gamma_{ret,1}$ = -.22 (.54), $t(309) = -.41$, $p > .05$.

***R-H2a*:** Choice deferral will negatively relate to FTP.
***R-H2b*:** Choice deferral will negatively relate to intended retirement.

Inspection of the latent path coefficient between latent variables Choice Deferral and FTP in the worker sample indicated a non-significant negative parameter estimate, γ_{52} = -.09 (.10), $t(309) = -.91$, $p > .05$. Inspection of the corresponding latent path in the non-retiree sample indicated a non-significant positive parameter estimate, γ_{52} = .08 (.26), $t(309) = .31$, $p > .05$.

Inspection of the structural path coefficient between latent variable Choice Deferral and measured variable Intended Retirement in the worker sample indicated a non-significant negative parameter estimate, $\gamma_{ret,2}$ = -.41 (.46), $t(309) = -.88$, $p > .06$. Inspection of the corresponding path coefficient in the non-retiree sample indicated a non-significant parameter estimate but, interestingly, in the opposite direction and contrary to the hypothesis, $\gamma_{ret,2}$ = 1.34 (.89), $t(309) = 1.56$, $p = .06$.

R-H3a**:** Decisional impulsivity will positively relate to FTP.

R-H3b**:** Decisional impulsivity will positively relate to intended retirement.

Inspection of the latent path coefficient between latent variables Decisional Impulsivity and FTP in the worker sample indicated a non-significant positive parameter estimate γ_{53} = .08 (.08), $t(309) = .99$, $p > .05$. Inspection of the corresponding latent path in the non-retiree sample indicated a similar non-significant positive parameter estimate, γ_{53} = .22 (.23), $t(309) = .96$, $p > .05$.

Inspection of the structural path coefficient between latent variable Decisional Impulsivity and measured variable Intended Retirement in the worker sample indicated a negligible parameter estimate, $\gamma_{ret,3}$ = .01 (.36), $t(309) = .69$, $p > .05$. Inspection of the corresponding path coefficient in the non-retiree sample also indicated a non-significant path coefficient, but in the opposite direction than hypothesized, $\gamma_{ret,3}$ = -.41 (.67), $t(309) = -.62$, $p > .05$.

R-H4a: Sunk costs will negatively relate to FTP.

R-H4b: Sunk costs will negatively relate to intended retirement.

Inspection of the latent path coefficient between latent variables Sunk Costs and FTP in the worker sample indicated a non-significant negative parameter estimate γ_{54} = -.15 (.08), t(309) = -1.88, p = .06. Inspection of the corresponding latent path in the older worker sample indicated a non-significant negative parameter estimate, γ_{54} = -.28 (.17), t(309) = -1.62, p = .06.

Inspection of the structural path coefficient between latent variable Sunk Costs and measured variable Intended Retirement in the worker sample indicated a non-significant positive parameter estimate, $\gamma_{ret,4}$ = .69 (.36), t(309) = 1.94, p = .05. Inspection of the corresponding path coefficient in the non-retiree sample indicated a non-significant positive parameter estimate, $\gamma_{ret,4}$ = -.72 (.52), t(309) = -1.39, p = .09.

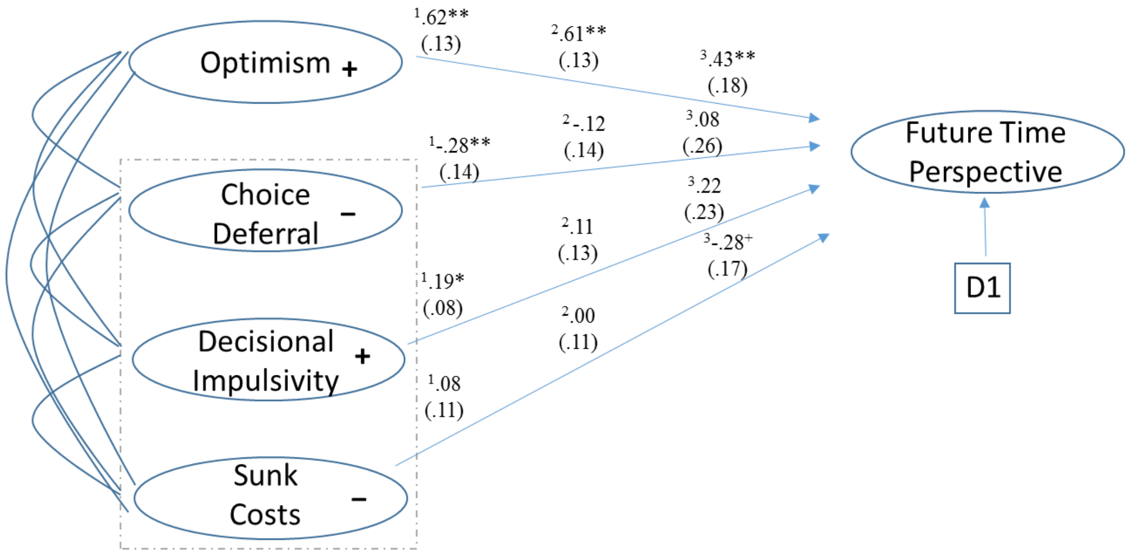


Figure 18. Summary of Obtained Main-Effect Structural Paramter Estimates for FTP for focal1, younger worker2, and non-retiree3 samples.46 +p < .10, *p < .05, **p < .01.

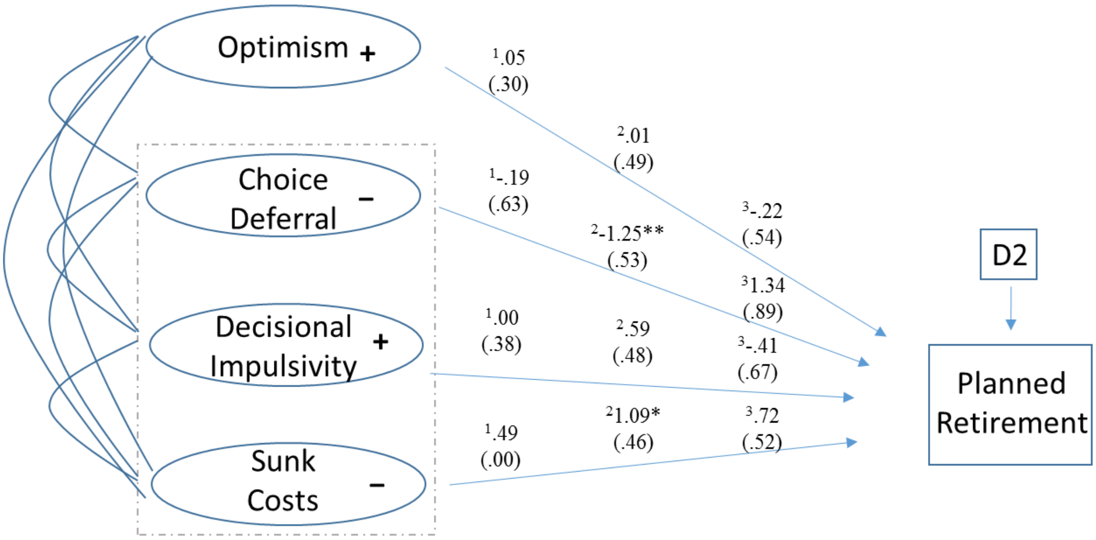


Figure 19. Summary of Obtained Main-Effect Structural Paramter Estimates for Planned Retirement for focal1, younger worker2, and non-retiree3 samples. +p < .10, *p < .05, **p < .01.

5.6 Retro-Observational Study Summary

The retro-observational study sought to examine hypotheses relating to the psychometric performance of the FTP questionnaire across target samples, also, to examine the nomological validity evidence of FTP within a latent-analytic framework. Findings from IRT analyses supported hypothes regarding higher discrimination and difficulty parameter estiamtes for FTP items among retirees,

---

[46] It should be noted that the depiction of individual outcomes in isolation is for display-pictoral clarity only. Analyses and reported obtained structural parameter coefficients (unstandardized) were conducted within a multivariate-analytic framework.

relative to workers of the same age-band. Extrapolating from these findings, it may be inferred that retirees are more precisely assessed, and are relatively higher on the theoretical underlying latent trait of FTP, relative to workers.

Results of the SEM analyses indicated some support for hypothesized predictors of FTP, with less support for the observed-variable, planned retirement. Specifically, three of four hypothesized main-effects were supported regarding the prediction of FTP in these data. No support was found, however, for the main-effect hypotheses regarding observable variable, intended retirement.

Contrast-sample analyses, however, helped depict potential sample-compositional effects regarding predictors of FTP and planned retirement. Specifically, the hypothesized negative main-effect of Chocie Deferral on intended retirement was found to be significant among younger workers. On the other hand, the corresponding negative main-effect for Sunk Costs exhibited a significant positive main-effect. It was speculated this was due to the retention of reverse-scored items for Sunk Costs from the measurement model specification, which may have resulted in a 'hardiness' factor. A subsequent parceling methodology corroborated this conjecture. An additional contrast-analyses revealed no appreciable differences in prediction equations for non-retirees, with the exception of the hypothesized negative prediction of Sunk Costs on FTP. This finding is concordant with the literature where Carstensen reports lower FTP as a function of life expectancy and situational constraints related to employment.

Moderation hypotheses were largely unsupported in these data. A parceling metholodology resulted in some support of the moderation hypotheses of SLE on predictors of FTP and intended retirement. Unsurprisingly, the contrast samples did not provide further info with regards to moderation hypotheses. Essentially, the additional contrast analyses were most informative for examimining the prediction of intended retirement, while parceling was most informative for examining moderation effects.

Future research on the measurement of FTP may consider a bi-factor model for measuring the latent space, otherwise, a multidimensional model in a non-linear framework, such as IRT would also be useful. Future research may also advance the methodological assessment of FTP by examining self-reports in conjuction with delay-discounting tasks. The next chapter present the experiment study.

CHAPTER 6

EXPERIMENTAL STUDY

6.1 Overview

The experimental study adopted two latent variable-analytic frameworks in order to examine sets of between-subject hypotheses and crossed-subject hypotheses. SEM analyses were conducted in order to examine between-subject hypotheses, including tests of main- and interaction-effects. Specifically, main-effect hypotheses pertained to the attribution framing (manipulation) effects on subjective life expectancy (SLE), in turn, FTP and intended retirement. The interaction hypotheses pertained to retirement planning's moderation effects vis-à-vis treatment main-effects. In addition, IRT analyses were conducted in order to examine cross-subject hypotheses, specifically, regarding the anchor-vignette rescaling procedure for strengthening personality scores as external correlates of FTP, only.

6.2 Methods

As with the retro-observational study, treatment of experimental data was conducted in SPSS v22.0. A total of N = 410 survey responses was recorded. Respondents who desisted from the survey following informed consent (N = 25) or the demographics section (N = 41) were identified as system-missing and subsequently removed from the dataset, resulting in N = 344. Participants identified as user-missing consisted primarily of attriters following the first substantive section of the survey (N = 48), resulting in N = 296 for formal hypothesis testing.

Two samples were collected for the experimental study: 1) A non-probabilistic, purposive community sample[47], and 2) A probabilistic, simple random sample. The different methods theoretically increased coverage of the intended target-population, however, statistical validity may be compromised between methods. Initial descriptive summary statistics indicated preliminary evidence for equivalence between the two samples, based on demographics and average-treatment effects on subjective life expectancy (SLE). In addition, a one-sample t-test on mean-SLE between concordant treatments indicate non-significance. The finding provided inference for the statistical equivalence between samples, $t_{Live}(116) = .10$, $p = .92$, $t_{Die}(117) = 1.09$, $p = .28$. The two samples were subsequently aggregated for tests of formal hypotheses. Summary descriptive statistics for each sample are reported in Table 29 below.

[47] Snowball, chain, or referral sampling are common informal labels, also, the sampling method is considered purposive rather than convenient, if the selection of distributors is strategic for accessing the target population. For the current sample, this consisted of aging research contacts with accessible participant lists and collegiate contacts in south Florida.

Table 25

Summary Descriptive Statistics by Community and Online Samples

Sample (*N*)	Variable	Mean (S.D.)	Variance	Range	Skewness	Kurtosis
	Duration	32.73 (36.19)	1309.10	190.27	2.96	9.85
	Age	54.44 (4.04)	16.30	[49 – 65]	.56	-.66
	Gender	1.51	-	-	-	-
Community	Prob Live 65	91.2 (1.12)	1.25	5	-1.64	3.25
(*N* = 86)	Prob Live 75	80.0 (1.54)	2.38	5	-0.29	-1.06
	Prob Live 85	58.8 (2.16)	4.68	8	-0.69	-0.49
	Prob Live 95	34.7 (2.24)	5.02	7	0.30	-1.27
	Prob Die 65	24.4 (2.07)	4.30	8	1.54	1.62
	Prob Die 75	41.2 (2.26)	5.11	9	0.65	-0.23
	Prob Die 85	64.4 (2.36)	5.59	8	0.02	-1.04
	Prob Die 95	80.0 (2.73)	7.43	9	-1.58	1.41

Sample (*N*)	Variable	Mean (S.D.)	Variance	Range	Skewness	Kurtosis
	Duration	12.66 (8.77)	1326.23	56.30	1.36	3.58
	Age	5.51 (4.17)	17.36	[49 – 65]	5.51	0.27
	Gender	1.50	-	-	-	-
Online	Prob Live 65	7.76	5.75	9	-1.23	0.73
(*N* = 230)	Prob Live 75	7.16	4.53	9	-1.10	0.72
	Prob Live 85	5.93	5.25	9	-0.28	-0.46
	Prob Live 95	4.44	7.79	9	0.35	-1.06
	Prob Die 65	3.57	6.02	8	0.55	-0.92
	Prob Die 75	5.12	5.13	9	-0.23	-0.87
	Prob Die 85	6.59	5.85	9	-0.65	-0.31
	Prob Die 95	7.67	7.43	9	-1.11	0.15

Note. Actual sample size varies slightly ($< N = 5$) for between-subject condition variables. Variable ‘duration’ is scaled in minutes. Prob = probabibility.

6.3 Results

6.3.1 Tests of Main-Effect Hypotheses

It was hypothesized that the attribute framing manipulation of subjective life expectancies (SLE) would lead to mean-level differences on the outcomes, future time perspective (FTP) and retirement intention. Preliminary evidence for the effectiveness of the manipulation indicated that the unweighted-mean difference (80-years old) was approximately $M = 1.02$ (.22), $t(235) = 4.581$, $p < .01$. That is, participants in the 'live-to' treatment reported, on average, an approximately 10.02% higher chance of living to age 80, relative to the 'die-by' condition.

Initial inspection of the estimated structural model indicated a well-fitting model to the observed sample data with robust estimates of, SB-$X^2(42, N = 178) = 53.62$, $p = .11$, AASR = .04, RMSEA = .03, CFI = .98. The distribution of standardized residuals also appeared approximately symmetric centered about zero. In order to test formal main-effect hypotheses, obtained structural parameter estimates corresponding to referenced latent paths were inspected. Results are reported below.

E-H1: The 'die-by' frame will decrease mean-FTP, relative to the control group.

E-H2: The 'die-by' frame will decrease intended retirement, relative to the control group.

Inspection of the latent path coefficient between the 'die-by' condition and latent variable FTP indicated a non-significant parameter estimate in the hypothesized direction, γ_{1D} = -.20 (.14), $t(55) = -1.39$, $p = .17$, providing no support for hypothesis E-H1.

Inspection of the latent path coefficient between the ‘die-by’ condition and latent variable Intended Retirement indicated a non-significant parameter estimate in the hypothesized direction, γ_{2D} = -.21 (.14), $t(55) = -1.48$, $p = .14$, providing no support for hypothesis E-H2.

E-H3: The ‘live-to’ frame will increase mean-FTP, relative to the control group.

E-H4: The ‘live-to’ frame will increase mean-intended retirement age, relative to the control group.

Inspection of the latent path coefficient between the ‘live-to’ condition and latent variable FTP indicated a significant parameter estimate in the hypothesized direction, γ_{1L} = 1.79 (.47), $t(55) = 3.84$, $p < .01$, providing support for hypothesis E-H3.

Inspection of the latent path coefficient between the ‘live-to’ condition and latent variable Intended Retirement indicated a significant parameter estimate in the hypothesized direction, γ_{2L} = 1.25 (.46), $t(55) = 2.71$, $p < .01$, providing support for hypothesis E-H4.

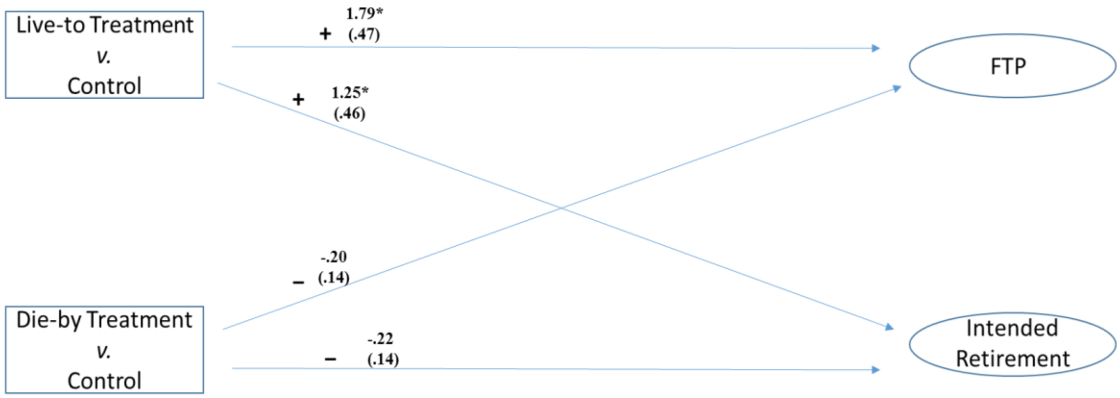


Figure 20. Unstandardized obtained latent-structural parameter estimates for experiment. ${}^{+}p < .10$, $*p < .05$, $**p < .01$.

6.3.2 Tests of Interaction-Effect Hypotheses

In order to test for potential interaction effects between latent variable Retirement Planning and the ‘live-to’ and ‘die-by’ experimental conditions on FTP and Intended Retirement, a similar multisample approach was adopted as that used in the retro-observation study. Specifically, low and high Retirement Planning groups were formed based on a median split (50%) on the proposed moderator. Raw data from the low-Retirement Planning (N = 129) and high-Retirement Planning (N = 129) groups were used to obtain estimates for the multiple-group SEM analysis. Equality constraints were then imposed on latent structural coefficients and decrement to global model fit assessed. Substantive noninvariance of latent structural paths as indicated by relative model misfit from the freely estimated structural paths in a multisample analysis is admitted as evidence of moderated effects as a function of levels of Retirement Planning. Summary descriptive statistics and intercorrelations matrices for low- and high-Retirement Planning groups on study variables is depicted below.

Initial inspection of the unconstrained multisample model with all parameter estimates freely estimated within- High- and Low-Retirement Planning -groups indicated a very well-fitting model to the observed sample data with robust estimates of, SB-X^2(84, N =278) = 54.88, p = .99, AASR = .04, CFI = 1.00. The distribution of standardized residuals also appeared approximately symmetric, centered about zero.

Next, freely estimated factor loadings within-groups were constrained to equality across groups and the model was reestimated. The reestimated model indicated a slight decrement to overall fit, but still fit the observed sample data very well overall, YB-X^2(91, N =278) = 85.93, p = .63, AASR = .04, CFI = 1.00. Next, freely estimated latent path coefficients within-groups were constrained to equality across and the model was reestimated. The restricted model, again, indicated a slight decrement to overall fit, but still exhibited overall good fit to the observed sample data, YB-X^2(95, N =278) = 189.97, $p < .00$, AASR = .04, CFI = 1.00. The change in chi-square values from unconstrained and constrained latent paths indicated significance $\Delta YB\text{-}X^2$ (4) = 104.04, $p < .00$. The finding provided omnibous evidence for moderation effects as a function of retirement planning.

Table 26

Intercorrelation Matrices and Summary Descriptive Statistics for Low- and High-Retirement Planning Groups

Low	FTP1	FTP2	FTP3	FTP4	RetInt1	RetInt2	RetInt3	RetInt4	RetInt5	***M***	***SD***
FTP1	1	.71	.59	.65	-.05	-.05	.05	.09	.09	17.32	4.64
FTP2	.71	1	.80	.84	.09	.07	.13	.16	.18	16.07	4.42
FTP3	.59	.80	1	.78	.12	.07	.12	.14	.18	15.71	5.20
FTP4	.65	.84	.78	1	.02	-.04	.07	.09	.01	14.22	5.04
RetInt1	-.05	.09	.12	.02	1	.74	.82	.77	.74	2.12	1.40
RetInt2	-.05	.07	.07	-.04	.74	1	.73	.75	.80	1.80	1.22
RetInt3	.05	.13	.12	.07	.82	.73	1	.97	.73	2.17	1.56
RetInt4	.09	.16	.14	.09	.77	.75	.97	1	.75	2.19	1.50
RetInt5	.09	.18	.18	.01	.74	.80	.73	.75	1	2.29	1.29

High	FTP1	FTP2	FTP3	FTP4	RetInt1	RetInt2	RetInt3	RetInt4	RetInt5	***M***	***SD***
FTP1	1	.692	.472	.581	.157	-.078	.053	.048	.091	18.42	4.50
FTP2	.692	1	.677	.788	.032	-.057	-.045	-.072	-.027	17.99	4.41
FTP3	.472	.677	1	.751	.020	-.067	-.089	-.087	-.017	16.90	4.76
FTP4	.581	.788	.751	1	.100	-.013	.014	-.004	.029	16.90	4.81
RetInt1	.157	.032	.020	.100	1	.441	.668	.665	.732	1.86	1.17
RetInt2	-.078	-.057	-.067	-.013	.441	1	.588	.661	.588	1.69	1.10
RetInt3	.053	-.045	-.089	.014	.668	.588	1	.822	.701	2.19	1.45
RetInt4	.048	-.072	-.087	-.004	.665	.661	.822	1	.779	2.16	1.33
RetInt5	.091	-.027	-.017	.029	.732	.588	.701	.779	1	2.40	1.43

Note. N = 129 for both Low and High groups.

Given the pattern of findings, moderation analysis commenced with the inspection of concordant structural parameter estimates across Low- and High-retirement Planning groups. Unstandardized structural parameter estimates were tested for significantly different magnitudes, and the effect of the group on the estimator was considered with the hypotheses relating Retirement Planning to the effects of the experimental conditions on FTP and Intended Retiement. The procedure was conducted in the simultaneously estimated population with constrained factor loadings, but freely estimated latent path coefficients. Figure 21 depicts the obtained latent path estimates across Low- and High- Retirement Planning groups, below.

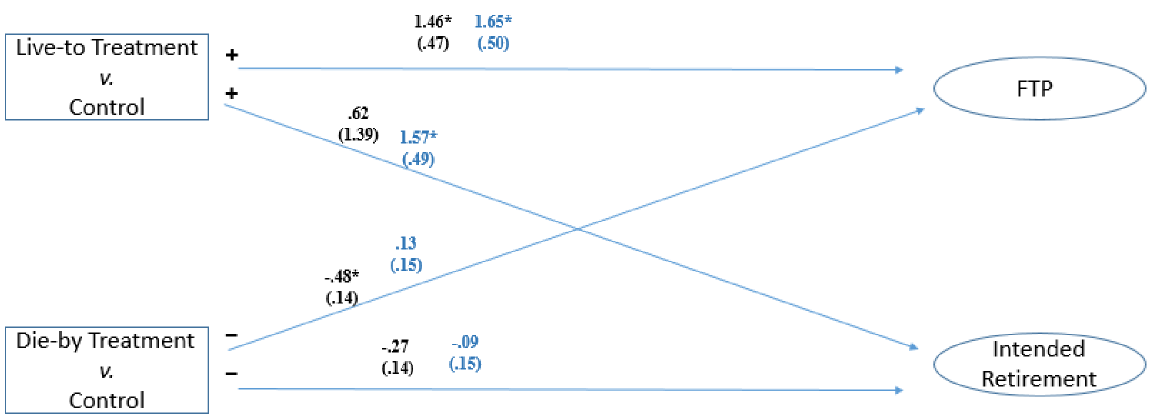


Figure 21. Unstandardized obtained latent-structural parameter estimates for experiment main-effects across low- and high-retirement planning groups. $^{+}p < .10$, *****$p < .05$, $**p < .01$. Black estimates are for the low-Retirement Planning group, blue estimates for High-retirement group.

E-H5: Retirement planning will interact with the 'die-by' frame on mean-FTP, such that more retirement planning will buffer the negative effect, relative to the control group.

Observing the non-significant, but positive latent path coefficient obtained for the High-Retirement Planning group, results provide tentative support for E-H5.

E-H6: Retirement planning will interact with the 'live-to' frame on mean-FTP, such that retirement planning will strengthen the positive effect, relative to the control group.

Observing the comparable effect size estimate obtained for the High-Retirement Planning group, results fail to support E-H6. The marginal increase in effect size was consistent with the hypothesis.

E-H7: Retirement planning will interact with the 'die-by' frame on intended retirement, such that retirement planning will strengthen the negative effect, relative to the control group.

Observing the attenuated negative latent path estimate from Low- to High-Retirement Planning groups, results fail to support hypothesis E-H7.

E-H8: Retirement planning will interact with the 'live-to' frame on intended retirement, such that retirement planning will reduce the positive effect, relative to the control group.

Observing the strengthened positive latent path estimate from Low- to High-Retirement Planning groups, results fail to support hypothesis E-H8.

The fial section of this chapter presents the empirical findings for covariance-structure hypotheses related to the within-subject experimental conditions.

6.3.3 'Pre – Post' Conditions and Big-Five Correlates with FTP

Summary descriptive statistics for composite outcomes and the Big-Five's personality factor indicators, by 'pre – post' conditions, is presented in Table below.

Independent-samples *t*-tests provide preliminary evidence for equality of means across conditions, with the exception of item $j_{\text{Disorganized}}$, an indicator of Big-Five factor Conscientiousness. Pertinent to the covariance-structure hypotheses, Levene's tests were also conducted on the composite outcomes and Big-Five's factor indicators, specifically, in order to provide preliminary inference on homoskedasticity assumptions (homogeneity of variance) across 'pre – post' conditions. Results indicated half of the total Big-Five indicators (j = 5) violated assumptions of homosekdasticity, which may have implications for instrument scaling and calibration. A latent-analytic model test of covariance-structure hypotheses is presented next.

Table 27

Summary Descriptive Statistics by Pre – Post Order

Variable	Order	*M*	*t*-value	SD	Levene's *W*
Future Time Perspective	Pre	65.23	-.54	17.59	.46
	Post	66.42		16.47	
Planned Retirement	Pre	10.23	-.30	6.33	.22
	Post	10.46		5.45	

Variable	Order	*M*	*t*-value	SD	Levene's *W*
Extraversion	Pre	4.60	.02	1.82	1.98
	Post	4.59		1.96	
Reserved	Pre	4.26	-1.36	1.81	.09
	Post	4.59		1.83	
Anxious	Pre	3.13	.88	1.89	1.30
	Post	2.92		1.71	
Calm	Pre	5.36	-1.27	1.59	8.30^{**}
	Post	5.60		1.20	
Open	Pre	5.25	-.13	1.35	2.78^{+}
	Post	5.27		1.57	
Conventional	Pre	3.06	-1.56	1.72	7.48^{**}
	Post	3.44		2.00	
Dependable	Pre	5.88	1.15	1.21	.10
	Post	5.71		1.14	
Disorganized	Pre	2.22	-2.22^{*}	1.36	11.28^{**}

	Post	2.66		1.65	
Sympathetic	Pre	5.47	-1.16	1.55	4.52*
	Post	5.68		1.23	
Critical	Pre	2.45	-1.05	1.57	.64
	Post	2.67		1.64	

Note. Independent-samples t-test. Levene's *W*-statistic follow *F*-distribution. N_{Pre} = 110, N_{Post} = 126. **$p < .01$, *$p < .05$, $^{+}$ <.10

A preliminary investigation of the impact, if any, of 'Order' effects between personality factors and FTP may include testing for equivalence of sample covariance matrices ($S_{Pre} = S_{Post}$). An inital model evaluated the similarity of personality factor structures (loadings) across the 'pre – post' conditions by imposing equality constraints on factor loadings across groups. Global model fit indices were evaluated to determine what, if any, impact assumptions of homogeneous sample covariance matrices would have on the multisample, global model fit indices.

Initial estimation of freely estimated factor loadings within-group indicated a fairly well-fitted model to the observed sample data, YB-X^2 (158, N = 236) = 180.46, p = .11, CFI = .89. Imposing equality constraints on previously freely estimated factor loadings led to a substantive decrement in global model fit indices, YB-X^2 (163, N = 236) = 198.23, p = .03, CFI = .87.. The finding provides evidence for factorial inequivalence of the Big-Five personality scores between 'pre' and 'post' measurement conditions.

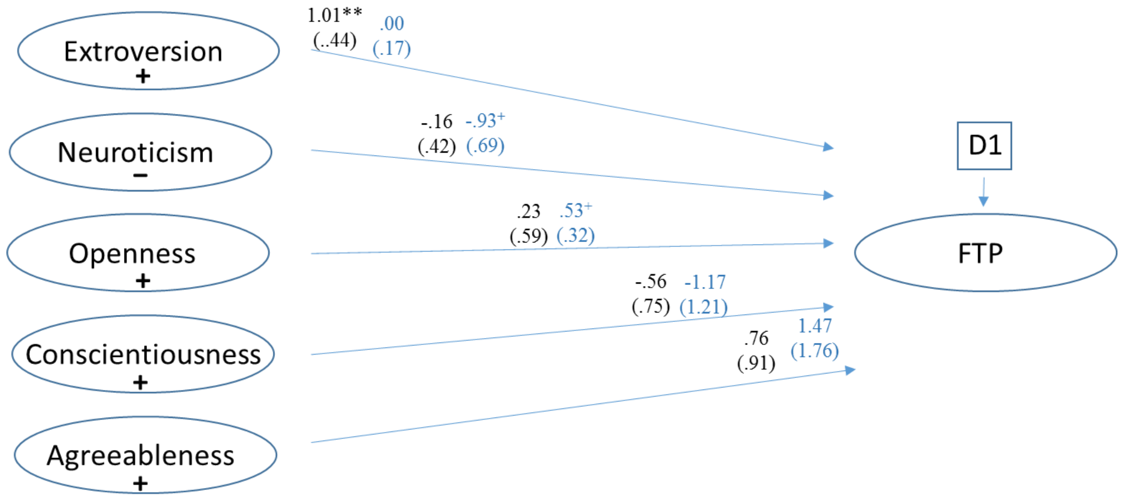


Figure 22. Unstandardized obtained latent-structural parameter estimates for ‘pre’ and ‘post’ conditions of personality factor correlates with FTP. $^{+}p < .10$, *****$p < .05$, $**p < .01$. ‘pre‘ condition estimates in black, ‘post‘ condition estimates in blue.

CHAPTER 7

GENERAL DISCUSSION

Current joint-standards emphasize systematic test development through recruitment of manifold evidence for test design and application (AERA, 2014). Two necessary sources of construct validity (CV) evidence for meaningful test use are internal (*construct representation*) and exernal (*nomothetic span*) (Embretson, 1983). They are generally conjunctive, so that information from one source of validity evidence should inform the body of evidence of another, ensuring explication of data – theory relations (see, Embretson, 2007).

The current dissertation identified a construct, Future Time Perspective (FTP), with disproportionate external validity evidence in a maturing population, specifically, workers and retirees. As social scientists from functionalist paradigms have begun to impugn the utility of chronological age as a proxy-variable over the lifespan, so too, have structuralists recommended diverse methods for increasing accuracy of test scores' meaning. The mono-operation bias[48] inherent totwoself-report instruments of FTP across functionally dissimilar populations (workers and retirees), thus, vesseled weak validity evidence and poor versimiilitude (Cronbach, 1988, Meehl, 1990).

A construct representation approach was adopted for this dissertation in order to examine the CV of FTP. Remphasizing internal sources of validity is argued to, not only advance theory-driven measurement, but also clarify sources of conflicting predictions *vis-à-vis* external correlates under the unifying concept of CV (Embretson, 2007). The

[48] Common-method bias (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003).

reaminder of the Discussion is organized into four subsections. First, obtained empirical findings from the dissertation are summarized, including limitations of the specific findings and more general paradigm of research. Second, theoretical and practical implications of the findings will be elaborated with emphasis on measurement issues. Third, substantive propositions are offered; they are guided by plausible rival hypotheses and aim to calrify extant descriptive organizational research (primarily, external validity evidenced) and stimulate further integration of the the work – retirement research domains (Campbell, 1960). Finally, a conclsucion is drawn from the current dissertation.

7.1 Summary of Empricial Findings

Chapter four consolidated empirical correlates of FTP's nomological network. Stronger evidence was found for Carstensen's (1994) instrument, relative to Hershey's (2000). The instruments were most distinguished by age with consistent, moderate to large negative correlates of Carstesen's instrument, and negligible relations with Hershey's.

Chapter five benchmarked the psychometric properties of Carstensen's FTP instrument using IRT analyses. Support was found for the two hypotheses of higher discrimination and location parameter estimates for the C-FTP items among retirees, relative to similarly aged workers. In addition, chapter five presented findngs from a latent-structural model of predictors of FTP and intended retirement. Support was found for three-of-four main-effect hypotheses regarding FTP, but little support for the same hypotheses for predicting the observed variable, intended retirement. Specifically, Optimism and Decisional Impulsivity emerged as positive significant predictors of FTP, while Choice Deferral exhibited negative predictive validity. Sunks Costs did not predict

FTP, however, contrast analyses illustrated its negative predictive validity (as hypothesized) in a non-retiree, unemployed subsample. The finding accords with the theoretical tenets of FTP in Socioemotional Selectivity Theory (Carstensen, 1993; SST) for the situational-sensitivity of FTP (out of work), also, with early empirical findings from Teahan and Kastenbaum (1970) in a population of long-term unemployed adults.

Contrast analyses, premised on focal sample selection factors (age, job status) helpedillustrate the nature of some main-effect relationships.Specifically, for FTP, Optimism exhibited stable positive prediction, while Choice Deferral and Decisional Impulsivity were both reduced to non-significance in a younger worker sample and an all-age non-retiree sample. Sunk Costs, however, indicated negative, significant predictive validity in the hypothesized direction for the non-retiree sample.

The contrast analyses also afforded better prediction of the observable variable intended retirement and, possibly, implicating range restriction on the observable outcome in the focal sample. In the younger worker sample, Choice Deferral emerged as a negative significant predictor of planned retirement. Sunk Costs also emerged as a significant predictor in the younger worker sample, but in a direction that was counter to the hypothesis. It was speculated that the finding may have been owed to the initial measurement model specification procedure, that is, items retained as indicators of latent factor Sunk Costs were uniformly reverse-scored and, potentially, tapped a psychological hardiness – resilience trait (Block & Block, 1980). Indirect support for this explanation came from subsequent parceling procedures for measurement model specification, whereby, all items were retained for Sunk Costs, and the corresponding obtained parameter estimate indicated non-significance, $\gamma = .11$ (.33), $t(215) = .32$, $p > .05$.

Moderation hypotheses were also tested regarding Subject Life Expectancy's (SLE's) effect on prediction equations. Little support was found for SLE's moderating effect across outcomes, however, a parceling method in the SEM framework proved useful for more fully evaluating SLE's potential moderating effects. Specifically,

Chapter six presented evidence from a '3 x 2' mixed-subjects experiment designed to test functional antecedents of FTP and Planned Retirement. Subjective Life Expectancy served as between-subjects factor via an attribute framing manipulation ('live-to', 'die-by', 'control') and Order ('pre', 'post') as within-subjects factor for administration sequence of the personality assessment battery. The Subjective Life expectancy manipulation was successfully replicated and hypotheses about main-effects were supported in terms of direction, but only reached significance in the 'live-to' condition. Specifically, participants assigned to the 'live-to' condition reported significantly higher mean-FTP and mean-Planned Retirement, compared to the control.

As expected, no mean-effect was found on either FTP or Planned Retirement composites as a function of the within-subject factor (Order). Big-Five personality factor scores also did not evidence significant mean-differences across 'pre' and 'post' conditions. Also, the moderation effects of Retirement Planning on the mean-level outcomes did not reach significance.

In addition to mean-level findings, the experiment also investigated covariance-structure hypotheses with regards to potential respose bias in self-reports. Covariance-structure analyses indicated stronger correlates in the hypothesized directions for four of the Big-Five factors to FTP, that is, the correlates were stronger in the 'post' condition relative to the 'pre' condition. Also, analyses indicated dissimilar latent-factor correlations of the Big-Five as a function of ‘pre’ and ‘post’ conditions. As expected, Big Five personal factor scores evidenced heterskeasticity (heterogeneity of variance) across ‘pre’ and ‘post’ conditions, which has implications for instrument calibration and scaling.

7.2.1 Practical Implications

Practical implications of this dissertation intersect the age-integration of social institutions and domain-integration of work-life balance. To this end, results should inform ongoing areas of research on related lifespan concepts, for example, active aging, aging-in-place, and successful aging in the workplace (Center for Disease Control and Prevention [CDC], 2011; Krause, 2001; World Health Organization [WHO], 2002,). Some practical utility becomes patently obvious with a cursory review of a particularly pressing challenge for policymakers with regards to social welfare systems design, namely, the underestimation of retirement by late-career employees. It is tenable that such underestimation could, partially, be a function of presentation of survey content. For example, if “retire-by” may be likened to “die-by”, and “live-to” may be likened to “work-to”, then cognitive processes engaged for survey responses may systematically vary.

Regarding workplace adaptations of FTP, redesigning work to motivate older employees’ may not function as increase of occupational-FTP, rather, as a *decrease* in waiting costs (Paglieri, 2013). In this sense, retirement planning should offset these effects, because those who have planned adequately for retirement will perceive a higher cost of waiting, i.e., delaying retirement a la’ motivational gradients. As Paglieri observes, “Costs of delay should be taken into account, on par with delay, in determining intertemporal choices. From this viewpoint, refusing to sustain the costs of a delay is not the same thing as devaluing a reward because it is delayed" (2013, p.371). Also regarding temporal frames, Mata & Hertwig, 2011 note, "it has become clear that motivational theories need to specify the reach of their predictions, in particular, whether there are age-specific changes that unfold at the anticipation, experience, and remembrance phases" (*p.373*)[49]. This has implications for questionnaire data, as well, such as cueing to specific time horizons when soliciting non-ability information.

7.2.2 Theoretical Implications

Generally, this author concurs with Shepard's (1993) articulation of consequential validity regarding unintended consequences of tests, specifically, that it is a "logical extension" (*p*.426) of Campbell's (1960) advocacy for rival hypotheses in validity evaluations. To this end, some examples of seemingly ambiguous and contradictory findings likely to lead to unintended consequences are identified below.

Many opportunities remain for the study of consequential forms of validity evidence within existing work and retirement domains. For example, a recent study that was thoughtfully designed revealed unintended consequences of data from measures of impulsivity and emotional intelligence (Winkel, Wyland, Shaffer, & Clason, 2011) in the work domain. Additional ambiguous and contradictory findings for related constructs within work and retirement research domains is elaborated below.

In another example of nomological ambiguity-breeding contradiction, in the organizational sciences, a popular heuristical conceptualization of factors relating to turnover are labeled 'push' and 'pull'. In a cursory review, these terms seem to first appear in an empirical study from Schultz (1998), but they lack any theoretical specificity

[49] This may be likened to the early notions of time "spheres" and phenomological frames.

or nomological network elaboration. These labels have endured, however, suffering from Dunnette's (1966) cautionary tale of "The Great Word Game" (*p.* 344). The issue in using these vague labels stems from their application across work and non-work domains. In contrast, the labels ('push' and 'pull') are restrictively ascribed negative valence (aversive stimuli) and positive valence (appetitive stimuli), respectively. Extrapolating, workers and retirees necessarily encounter push and pull factors, but they are functionally different across domain; Being pushed into retirement (age stereotypes) is likely unrelated to being 'pushed' to keep working (delay retirement). Conversely, the pull toward retirement (leisure) is likely unrelated to the pull to continue working (occupational identity). This author recommends abandonment of these heuristic labels, as they have never been systematically developed nor examined and stand to muddle the formidably complex domains of work and retirement (also, see Vancouver and Weinhardt, 2012).

In addition, pertaining to selection system design experts, the ability to identify criteria (ATIC) has been conceptualized as an individual difference construct in organizational behavior research (König, Melchers, Kleinmann, Richter, & Klehe, 2007). It is implicated as an explanatory variable for the validity of selection system devices, e.g., personality inventories, structured interviews. The eschewing of score use is shared by test developers and users, alike, and this likely comes at detriment to furthering the understanding between persons, the stimuli they encounter, and the precision with which their interactions can be validly assessed. When rigidity of test specifications (meaning) meets overreliance on psychometric properties of validity (significance), contradictions in hypotheses become probable. Further theoretical work in the organizational science may, for instance, advance along similar lines as that of developmental psychologists, abridging early childhood development and gerontological study (i.e., labor force entry and exit). Alternatively, organizational psychology theory would likely benefit from new perspectives on the meaning and purpose of work.

7.2.3 Limitations

The current dissertation proposal was motivated by conflicting predictions stemming from two psychological paradigms on future time perspective, however, there is programmatic research on FTP in educational assessment, as well (Nuttin, 1964; 1973). Unfortunately, the disintegration of the educational and psychological assessment paradigms vis-à-vis FTP is vestigial of the initial schism between developmental psychology into gerontology and early childhood development. It may be tenable, however, that common goals in terms of measurement between the two may be mutualistic for theoretical development. More formalized theoretic propositions aimed to stimulate future integration of work and retirement phenomena are provided in Appendix H.

Current test standards of integrated construct validity entail different sources of evidence that necessarily inform ongoing, multifaceted CV procedures (AERA, 1999, 2014; Cronbach & Meehl, 1955). New applications of measures should include internal sources of validity evidence. That is, it has long been recognized that overreliance on external sources of validity evidence may impugn the meaningfulness of test scores (Bechtoldt, 1959; Embretson, 1983). Construct representation is one aspect to validation efforts that prioritizes internal sources of validity evidence, thereby, repurposing nomological networks (external) from defining construct meaning to indicating the utility or significance of the test. Conversely, weak nomological validity evidence may also be indicative of, imprecise theory, deficient test development, or both. That is, both the target test and the tests which are used for external validation are impacted by test design and item design features. In particular, self-report data is known to be vulnerable to a number of cognitive and perceptual biases (Stone, 2000). Contemporary cognitive theory and modern test theory (item response theory [IRT]) are cornerstones to construct representation research for validation (Embretson, 1983; 1998).

This dissertation applied principles from cognitive theory to evaluate the effects of instruction manipulation (experimental) and response rescaling (self-report bias correction) on the psychometric properties and external correlates of self-reports. The dissertation adopts a case study of practical importance, namely, future time perspective (FTP) in work and retirement literatures, as a demonstrable application of construct representation research in CV. The evaluative experiment seeks to resolve opposing predictions for the developmental function (age-related correlates) of FTP across work and retirement domains of research. In addition, a rescaling procedure using personality anchoring vignettes to reduce bias in self-report data will be examined for its effects on, both the levels of FTP and its external correlates (King, Murray, Salomon, & Tandon, 2004).

Complementing the integral framework of CV, the dissertation first detailed a retrospective observational study of FTP. This study will employ latent variable modeling to examine evidence for the potential moderating effects of subject life expectancies (SLEs) on FTP and its correlates, which is pertinent to the manipulation in the subsequently proposed experiments. IRT analyses will also address potential measurement bias of the FTP scale and its comprising items as a function of the proposed manipulation. As Ferber observed, "the problem of response bias must be considered with specific reference to a particular question or characteristics" (Ferber, 1949, *p.* 672).

Theoretical implications will direct attention to greater synthesis of research across work and retirement domains, as well as advantages that construct representation research may offer toward integration. Methodological aspects of the dissertation necessarily, span theoretic and applied implications under the integrated framework of CV (Schwarz, 1999). In particular, the measurement procedures implemented in the current study will be elaborated with respect to comprehending self-report processes in data. To this end, a compendium of recent advances in measurement and modeling of self-report data will be overviewed to encourage future utilization.

7.3 Testable Propositions for Future Integration of Work – Retirement Scholarship

In terms of consequential validity, a recent adaptation of the C-FTP scale for application in the occupations domain need-be evaluated against potential, unintended consequences (Zacher. 2009; Zacher & Freese, 2010). For example, Zacher and colleagues argue that, by enhancing certain job characteristics intrinsically more appealing to older workers, retention may be increased by way of *increasing* their occupational FTP. Based on evidence from the current dissertation, a plausible rival hypotheses might suggest that enhancing valued job characteristics may increase retention, not by increasing O-FTP, but perhaps by reducing the psychological *costs* of delaying retirement. For example, Paglieri (2013) reviews the experimental literature on intertemporal choices indicating that waiting (psychophysical behavioral) versus postponing (choice-based delay) have markedly different effects on delay discounting (effect size difference by a factor of approximately 10). Therefore, the implications for older workers suggest that, the increased attractiveness of delaying retirement may be potentiated by inadequate retirement planning. This alternative hypothesis can be fairly easily evaluated in an experimental paradigm by factorially crossing self-report retirement planning, intended retirement, and occupational FTP, with two version of a delay discounting task – one based on hypothetical choice, one based on deferral of gratification (patience).

With regards to the impact of retirement planning on retiree wellbeing, two competing predictions are made, particularly with regards to the *transition* to (early phase of) retirement. Specifically, retirement planning has the greatest positive effect on wellbeing in early phases of retirement (Taylor & Doverspike, 2003; Taylor & Schaffer, 2012), whereas job-level is predicted to have the greatest *negative* impact in early phases. Planning may exhibit, either a preparatory / buffering effect or, otherwise, an expectancy effect. In an observational paradigm, then, *controlling for the 'voluntariness' of the retirement decision*, job-level should strengthen the positive effect of planning (in case of preparatory mechanism) on wellbeing. On the other hand, job-level should attenuate the effect of planning (in case of expectancy mechanism) on wellbeing.

In terms of career maturation and development, continuity should afford more resource-accumulation, whereas discontinuity may afford more *experience* with non-work time, potentiating adaptation and coping. As with the above proposition, these competing predictions may be tested in an observational study design, but should include a variable for voluntariness of the retirement decision.

7.4 Conclusion

Both the continued aging of the general population and increasing life expectancies necessitates precise application of lifespan theories of motivation to career development and, encompassing, to work – life domains. This dissertation presented an evaluative experiment designed to pit two competing hypotheses. Results provided some clarity on the developmental trend of future time perspective, but also provide an explanatory model for the current, dominant view of retirement as a timely or delayed event. Future research directions include, for example, a more diverse battery of measurements, including delay discounting (Bidwell, Griffin, & Hesketh, 2006) as, perhaps, a psychophysical indicator of future time perspective or retirement timing.

In the same manner that nomothetic span evidence can indicate the quality of construct representation in test construction, conversely, psychometric test data of, purportedly, the same measured construct may bear on utility from application of derived scores (Embretson, 1983). For example, when one considers the "famously ambiguous" criteria of retirement, the implications are evident (Ekerdt, 2010; p.70)[50]. In this sense, the divide between work and retirement scholarship may be an ecological case of disintegrated sources of validity evidence for continued CV research. In result, measures under the same label and premised on similar lines of rationale make contradictory predictions within their nomological network. As Cronbach and Meehl (1951) observed, "Rationalization is not construct validation." (p.291).

---

[50] Historically, Donahue, Orbach, & Pollak (1960) ascribe to retirement, "a certain degree of vagueness and lack of clarity" (p.330). Contemporary scholars opine the "ever-changing meanings of retirement" (McVittie & Goodall, 2012, p.75). Operationally, Denton & Spencer (2009) identify no-less than eight criterion definitions from a cursory review of the literature.

The researcher who endeavor to investigate CV under the unified model may be likened to the practitioner who necessarily declares hortatory and minatory service standards (Dawes, 1994). The criticism leveled against defining content validity in operational, rather than theoretical, terms is parity to interpreting external correlates as theoretical indices to the neglect of score meaning. As Janssen notes, “The difference between an item groups design and an item features design is merely at a conceptual level” (2010; *p.*231). More explicit, Janssen summarizes, “In domain-referenced testing, the principle idea is that the items of a test are a random sample from a domain, which refers to a population of items. Consequently, the item parameters of the test can be seen as random-effect parameters” (2010; *p.* 237). The current dissertation sought to combine and examine evidence from –seemingly- disparate sources in order to advance substantive paradigms in a more coherent direction.

APPENDIX A

RAND ALP DATA MANUAL STATEMENT ON PANEL SAMPLE WEIGHTING

2.9. Weights

As with all surveys based on random samples, the composition of the un-weighted ALP sample differs from the population composition. The ALP constructs sampling weights to correct for this sampling error and to make the sample as representative of the population of interest as possible. The benchmark distributions against which the ALP is weighted are derived from the Current Population Survey (CPS). This choice follows common practice in surveys of consumers, for example, the HRS. The sampling weights for the 2008 SCPC were constructed using the March 2008-2010 waves of the CPS, which includes the annual income supplement.

Three weighting methods have been implemented for the ALP: cell-based post stratification, logistic regression, and raking. After some experimentation, raking was found to give the best results among these different methods. It allows finer categorizations of variables of interest (in particular, age) than cell-based post-stratification does, while still matching these distributions exactly. Variables were created that account for interactions with gender, so that all distributions are matched separately for males and females. The resulting set of variables whose distributions are matched exactly is:

Gender × age, with 14 categories: (1) male, 18-24; (2) male, 25-34; (3) male, 35-44; (4) male, 45-54; (5) male, 55-64; (6) male, 65-74; (7) male, 75+. Categories (8)-(14) are the same as (1)-(7), except that they are for females instead of males.

Gender × race/ethnicity, with 6 categories: (1) male, non-Hispanic white; (2) male, non-Hispanic African American; (3) male, Hispanic and other; (4) female, non-Hispanic white; (5) female, non-Hispanic African American; (6) female, Hispanic and other.

Gender × (household) income, with eight categories: (1) male, <$25,000; (2) male, $25,000-$49,999; (3) male, $50,000-$74,999; (4) male, $75,000+; (5) female, <$25,000; (6) female, $25,000-$49,999; (7) female, $50,000-$74,999; (8) female, $75,000+.

Gender × education, with six categories: (1) male, high school or less; (2) male, some college or a bachelor's degree; (3) male, more than a bachelor's degree; (4) female, high school or less; (5) female, some college or a bachelor's degree; (6) female, more than a bachelor's degree.

All aggregate U.S. statistics for the SCPC were weighted using the sampling weights constructed in this manner.

Weights are currently provided on a per-request basis. In order to obtain weights for particular ALP data set, please contact the ALP at mmic@rand.org

APPENDIX B

ANALOGOUS MEASUREMENT MODEL TABLES (16 - 20), FIGURE (16), AND DESCRIPTIVE INTERPRETATION FOR BASE SAMPLE (*N* = 434)

Estimation results of the base 5-factor measurement model indicated poor global and local fit to the observed sample data, S-Bχ^2= 2221.35 (687, 432), $p < .00$; CFI = .77, AASR = .05, RMSEA = .07 [90% CI = .07, .08], see Table 7-B). Inspection of the standardized residuals and pattern factor loadings indicated localized areas of solution strain. Following respecification, the base 5-factor measurement model was reestimated and results indicated relatively improved fit to the observed sample data, S-Bχ^2= 859.19 (314, 432), $p < .00$; Y-Bχ^2 = 535.96 (314, 432), $p = .06$; CFI = .89, AASR = .04, RMSEA = .06 [90% CI = .06, .07]). Inspection of the standardized residuals and pattern factor loadings indicated localized areas of solution strain, specifically with the first item of the fourth factor (Sunk Cost) indicating a weak loading. Following respecification, the base 5-factor measurement model was reestimated and results indicated only a chi-square statistical improved fit to the observed sample data, S-Bχ^2= 814.28 (289, 432), $p < .00$; Y-Bχ^2 = 335.30 (289, 432), $p = .03$; CFI = .89, AASR = .04, RMSEA = .06 [90% CI = .06, .07]). Inspection of the standardized residuals and pattern factor loadings indicated adequate loadings of retained indicators, with the largest standardized residual contributing to misfit estimated between concatenate items at.254, compared to .571 in the base measurement model. It is noteworthy that the final measurement model solution for the conjectured 5-factor structure replicated items retained for the concordant solution in the listwise sample, although, in less rounds with the greater number of participants.

Table 16-Base.

Summary of Restricted Backward-search Model Respecification for the Base Sample (N = 434)

Round	**Iter**	**Factor Label**	**# *J***	**$\Lambda_{j,k} < .32$**	**$<\Lambda_k^2>$**	**Item Content**
1	9	Impulsivity	4	$\Lambda_{5,3} = .28$	.61	When making decisions, I do what seems natural at the moment.
		Sunk Costs	10	$\Lambda_{1,4} = .17$	-	I would waste a lot of time worrying about it instead of just doing something about it.
		Sunk Costs	9	$\Lambda_{2,4} = .31$	-	It would take my a long time to adjust myself to it.
		Sunk Costs	8	$\Lambda_{3,4} = .20$	-	I would feel paralyzed.
		Sunk Costs	7	$\Lambda_{4,4} = .24$	-	I would have trouble doing anything at all.
		Sunk Costs	6	$\Lambda_{5,4} = .15$	-	I wouldn't know how to deal.
		Sunk Costs	5	$\Lambda_{6,4} = .08$	-	I wouldn't have difficulty starting.
		Sunk Costs	4	$\Lambda_{7,4} = .29$	.46	I would take immediate action to correct it.
		FTP	11	$\Lambda_{10,5} = .16$	-	I have the sense that time is running out.
		FTP	10	$\Lambda_{11,5} = .09$	-	As I get older, I begin to experience time as limited.
		FTP	9	$\Lambda_{12,5} = .01$	-	I feel the importance of time.
		FTP	8	$\Lambda_{7,5} = .09$	.51	I have limited time left to live my life.
2	5	Sunk Costs	3	$\Lambda_{1,4} = .22$	.55	I would take action rather than just complaining about the situation.
3	4	-	-	-	-	

Note. Iter = # of Iterations for Optimal Parameter Estimate Convergence. # *J* = remaining indicators following respecification. **$< \Lambda_k^2 >$ =** the expected average communality index from remaining indicators.

Table 17-Base.

Nested-Measurement Model Comparisons for Base Sample (N =434)

Model	**S-B X^2**	***Df***	**CFI**	**AASR**	**BIC**	**RMSEA [90% CI]**	**Change Indicators**			
							[1]**ΔS-B X^2 / Δ*Df***	**ΔCFI**	**Δ*BIC***	**Δ*RMSEA***
1-Factor	3340.89	299	.39	.11	3656.69	.15 [.15 - .16]				
2-Factor	2714.98	298	.51	.11	3036.85	.14 [.13 - .14]	61.67 / 1*	.00	619.84	.01
3-Factor	1899.82	296	.68	.07	2233.84	.11 [.11 - .12]	138.63 / 2*	.04	803.01	.03
4-Factor	1180.01	293	.82	.06	1532.24	.08 [.08 - .09]	511.35* / 3	.14	701.60	.03
5-Factor	849.41	289	.89	.04	1225.94	.07 [.06 - .07]	226.14* / 4	.14	306.30	.01
6-Factor	804.89	284	.89	.04	1211.78	.07 [.06 - .07]	48.18* / 5	.00	14.16	.00
7-Factor	797.14	278	.89	.04	1240.47	.07 [.06 - .07]	8.75 / 6	.00	+28.69	.00

Note. * $p < .05$. 1. Scaled S-B chi-squared differences computed. CI = confidence interval.

Following latent factor identification, restricted forward specification search proceeded with inspection of the multivariate-Lagrange Multiplier (M-LM) modification index for potential improved fit to the measurement model (Chou & Bentle, 1990). The M-LM statistics asymptotically approach a chi-square distribution (Bentler & Dijkstra, 1985). The simultaneous procedure was implemented.

With the base 5-factor model, inspection of the M-LM index indicated free estimation of two observed variable error-covariances could improve the model's fit. The statistical cues of improved local-fit were subsequently probed with three substantive deliberations for free estimation consideration: 1) Common-factor loadings of corresponding congeneric indicators (Hooper, 2008), 2) Methodological justification within context of questionnaire responding (Brown & Moore, 2002), and 3) Theoretical soundness of homogeneity for item-content domain (Anderson & Gerbing, 1982). Application of the decision-criteria permitted the two error-covariance terms to be freely estimated. The modified model was then reestimated and, again, the multivariate M-LM modification index indicated potential improved fit with two additional freely estimated error covariances. Following the same decision criteria, the two errors were permitted to covary and the model was reestimated (Chou & Bentler, 1996). The modified measurement model was reexamined against a one-factor and six-factor solution, supporting the retention of the 5-factor model (Table 3-B). The standardized solution for the base sample measurement model is displayed in Figure 1-B, below.

Table 18-Base.

Measurement Model Restricted-Forward Search Respecification Criteria for Base Sample (*N* =434)

						Substance		
Round	**M-LM Step**	**θ**	**$X^2_{(Inc)}$**	***p-val***	**Corresponding Indicator Content**	**$\Lambda_{.m}$**	**Meth**	**Thry**
1	1	E6,E5	126.61	<.00	I hardly ever expect things to go my way, I rarely count on good things happening to me.	x	c, r	x
	2	E5,E4	99.39	<.00	I hardly ever expect things to go my way, If something can go wrong for me, it will.	x	c, r	x
2	1	E23,E22	89.61	<.00	Most of my life lies ahead of me, My future seems infinite to me.	x	c	x
	2	E8,E7	53.03	<.00	I postpone decision-making whenever possible, I avoid making decisions until the pressure is on..	x	c	x

Note. * *p* < .00. 1. Scaled S-B chi-squared differences computed. CI = confidence interval. **$\Lambda_{.m}$ =** common-factor of corresponding indicators. $\hat{\theta}$ = parameter estimated for free estimation. $X^2_{(Inc)}$ = univariate increment of modification. Meth = methodological rationale, Thry = item-feature similarity for ordering on theoretical latent factor. x = generic satisfaction of substantive criterion, c = concatenate, r = reverse-score.

Table 19-Base.

Measurement Model Modification Fit Indices and Configural Comparisons for Base Sample (N = 434)

Model	**S-B X^2**	***Df***	**CFI**	**AASR**	**BIC**	**RMSEA [90% CI]**	**Change Indicators**			
Base	849.41	289	.89	.04	1225.94	.06 [.06 - .07]	**[1]LR-$X^2_{(1)}$**	**ΔCFI**	**Δ*BIC***	**Δ*RMSEA***
2-E Cov	688.16	287	.92	.04	1076.83	.06 [.05 - .06]	225.85 / 2*	.03	149.11	.0
4-E Cov	562.31	285	.94	.04	938.84	.05 [.04 - .05]	142.65 / 2*	.02	137.99	.01

Model	**S-B X^2**	***Df***	**CFI**	**AASR**	**BIC**	**RMSEA [90% CI]**	**Change Indicators** ΔS-B X² / ΔDf	**ΔCFI**	**ΔBIC**	**ΔRMSEA**
4-E 1 Fac	2862.67	295	.48	.10	3202.76	.14 [.14 - .15]	1210.20 / 10*			
4-E 4 Fac	889.78	288	.88	.05	1272..38	.07 [.06 - .08]	139.53 / 3*	-.06	95.88	.02
4-E 6 Fac	519.20	280	.95	.04	950.383	.04 [.04 - .05]	45.76 / 3*	.01	54.16	.01

Note. * p < .00. 1. Scaled S-B chi-squared differences computed. CI = confidence interval. Base = 5-factor model solution.

Table 20-Base.

Bivariate Correlation Matrix and Summary Descriptive Statistics for the Base Sample (*N* = 434)

	Opt	Opt	Opt	Opt	Opt	Opt	Def	Def	Def	Def	Def	Imp	Imp	Imp	Imp	SC1	SC2	SC3	FTP1	FTP2	FTP3	FTP4	FTP5	FTP6	FTP7	FTP8	**Sk**	**Krt**	**Ra**
Opt	1	.67	.65	.30	.44	.43	-.12	-.13	-.08	-.10	-.15	-.04	-.02	.06	-.09	-.20	-.16	-.12	.24	.22	.26	.19	.23	.19	.18	.26	-1.03	0.58	5
Opt	.67	1	.68	.37	.51	.45	-.14	-.17	-.10	-.08	-.18	-.08	-.03	.00	-.12	-.17	-.16	-.17	.26	.27	.30	.22	.28	.25	.19	.25	-0.69	-0.03	5
Opt	.65	.68	1	.43	.57	.56	-.10	-.08	-.07	-.05	-.11	-.02	.02	.03	-.03	-.18	-.16	-.14	.30	.25	.31	.19	.21	.16	.20	.18	-1.07	0.77	5
Opt	.30	.37	.43	1	.62	.46	-.17	-.15	-.13	-.11	-.13	-.02	.01	.07	-.04	-.02	-.07	-.09	.25	.22	.29	.09	.21	.15	.15	.23	-0.65	-0.69	5
Opt	.44	.51	.57	.62	1	.74	-.19	-.16	-.10	-.12	-.15	-.13	-.07	-.02	-.14	-.09	-.07	-.04	.22	.23	.25	.11	.14	.12	.15	.23	-1.02	0.37	5
Opt	.43	.45	.56	.46	.74	1	-.14	-.16	-.10	-.15	-.14	-.10	-.04	-.01	-.07	-.15	-.08	-.05	.22	.24	.25	.14	.16	.13	.21	.23	-1.07	0.28	5
Def	-.12	-.14	-.10	-.17	-.19	-.14	1	.63	.50	.52	.48	.17	.17	-.06	.22	.19	.16	.11	-.06	-.11	-.09	-.12	-.11	-.06	-.09	-.15	0.79	-0.01	4
Def	-.13	-.17	-.08	-.15	-.16	-.16	.63	1	.52	.50	.55	.23	.20	-.02	.20	.18	.21	.10	-.07	-.07	-.07	-.11	-.07	-.10	-.06	-.14	0.97	0.26	4
Def	-.08	-.10	-.07	-.13	-.10	-.10	.50	.52	1	.69	.59	.21	.20	.04	.23	.19	.22	.15	-.06	-.10	-.07	-.16	-.16	-.10	-.07	-.15	0.46	-0.66	4
Def4	-.10	-.08	-.05	-.11	-.12	-.15	.52	.50	.69	1	.59	.37	.32	.12	.29	.16	.17	.12	-.02	-.11	-.06	-.09	-.09	-.03	-.05	-.11	0.76	-0.25	4
Def	-.15	-.18	-.11	-.13	-.15	-.14	.48	.55	.59	.59	1	.23	.18	.05	.29	.25	.24	.17	-.07	-.11	-.09	-.13	-.12	-.18	-.13	-.17	1.01	0.33	4
Imp	-.04	-.08	-.02	-.02	-.13	-.10	.17	.23	.21	.37	.23	1	.78	.55	.55	.05	.03	.01	.00	-.03	-.04	.01	.02	.06	-.02	-.15	0.60	-0.29	4
Imp	-.02	-.03	.02	.01	-.07	-.04	.17	.20	.20	.32	.18	.78	1	.59	.64	.03	.00	-.05	.05	-.02	.02	-.02	.06	.05	.02	-.09	0.42	-0.56	4
Imp	.06	.00	.03	.07	-.02	-.01	-.06	-.02	.04	.12	.05	.55	.59	1	.49	-.06	-.09	-.13	.08	.02	.03	-.03	.03	.07	.01	-.06	-0.02	-0.73	4
Imp	-.09	-.12	-.03	-.04	-.14	-.07	.22	.20	.23	.29	.29	.55	.64	.49	1	.10	.07	.06	-.04	-.12	-.05	-.12	.00	-.02	-.04	-.16	0.56	-0.31	4
SC1	-.20	-.17	-.18	-.02	-.09	-.15	.19	.18	.19	.16	.25	.05	.03	-.06	.10	1	.63	.47	-.13	-.07	-.14	-.12	-.15	-.14	-.16	-.10	0.90	0.25	4
SC2	-.16	-.16	-.16	-.07	-.07	-.08	.16	.21	.22	.17	.24	.03	.00	-.09	.07	.63	1	.54	-.07	-.04	-.09	-.07	-.15	-.06	-.09	-.07	0.31	-0.46	4
SC3	-.12	-.17	-.14	-.09	-.04	-.05	.11	.10	.15	.12	.17	.01	-.05	-.13	.06	.47	.54	1	-.11	-.07	-.17	-.08	-.13	-.14	-.11	-.07	0.49	-0.55	4
FTP	.24	.26	.30	.25	.22	.22	-.06	-.07	-.06	-.02	-.07	.00	.05	.08	-.04	-.13	-.07	-.11	1	.66	.76	.48	.46	.50	.48	.55	-0.32	-0.62	6
FTP	.22	.27	.25	.22	.23	.24	-.11	-.07	-.10	-.11	-.11	-.03	-.02	.02	-.12	-.07	-.04	-.07	.66	1	.74	.46	.39	.51	.53	.44	-0.19	-0.69	6
FTP	.26	.30	.31	.29	.25	.25	-.09	-.07	-.07	-.06	-.09	-.04	.02	.03	-.05	-.14	-.09	-.17	.76	.74	1	.50	.46	.56	.49	.49	-0.48	-0.34	6
FTP	.19	.22	.19	.09	.11	.14	-.12	-.11	-.16	-.09	-.13	.01	-.02	-.03	-.12	-.12	-.07	-.08	.48	.46	.50	1	.65	.51	.46	.38	0.30	-0.52	6
FTP	.23	.28	.21	.21	.14	.16	-.11	-.07	-.16	-.09	-.12	.02	.06	.03	.00	-.15	-.15	-.13	.46	.39	.46	.65	1	.46	.45	.43	0.25	-0.80	6
FTP	.19	.25	.16	.15	.12	.13	-.06	-.10	-.10	-.03	-.18	.06	.05	.07	-.02	-.14	-.06	-.14	.50	.51	.56	.51	.46	1	.61	.40	0.02	-0.94	6
FTP	.18	.19	.20	.15	.15	.21	-.09	-.06	-.07	-.05	-.13	-.02	.02	.01	-.04	-.16	-.09	-.11	.48	.53	.49	.46	.45	0.61	1	.39	0.07	-0.80	6
FTP	.26	.25	.18	.23	.23	.23	-.15	-.14	-.15	-.11	-.17	-.15	-.09	-.06	-.16	-.10	-.07	-.07	.55	.44	.49	.38	.43	0.40	0.39	1	-0.45	-0.84	6
SD	1.25	1.23	1.20	1.50	1.27	1.39	1.01	1.01	1.11	1.05	1.03	1.00	1.02	1.07	.99	1.02	1.07	1.18	1.53	1.57	1.50	1.60	1.75	1.66	1.56	1.65			
M	4.62	4.26	4.81	4.48	4.83	4.82	2.08	1.90	2.33	2.09	1.94	2.12	2.26	2.71	2.16	2.07	2.56	2.55	4.90	4.61	5.07	3.66	3.79	4.27	4.28	4.92			

Note. Opt = Optimism, Def = Choice Deferral, Imp = Impulsivity, SC = Sunk Costs, RtPln = Retirement Plan, Skw = Skew, Krt = Kurtosis, Ran =

Range.

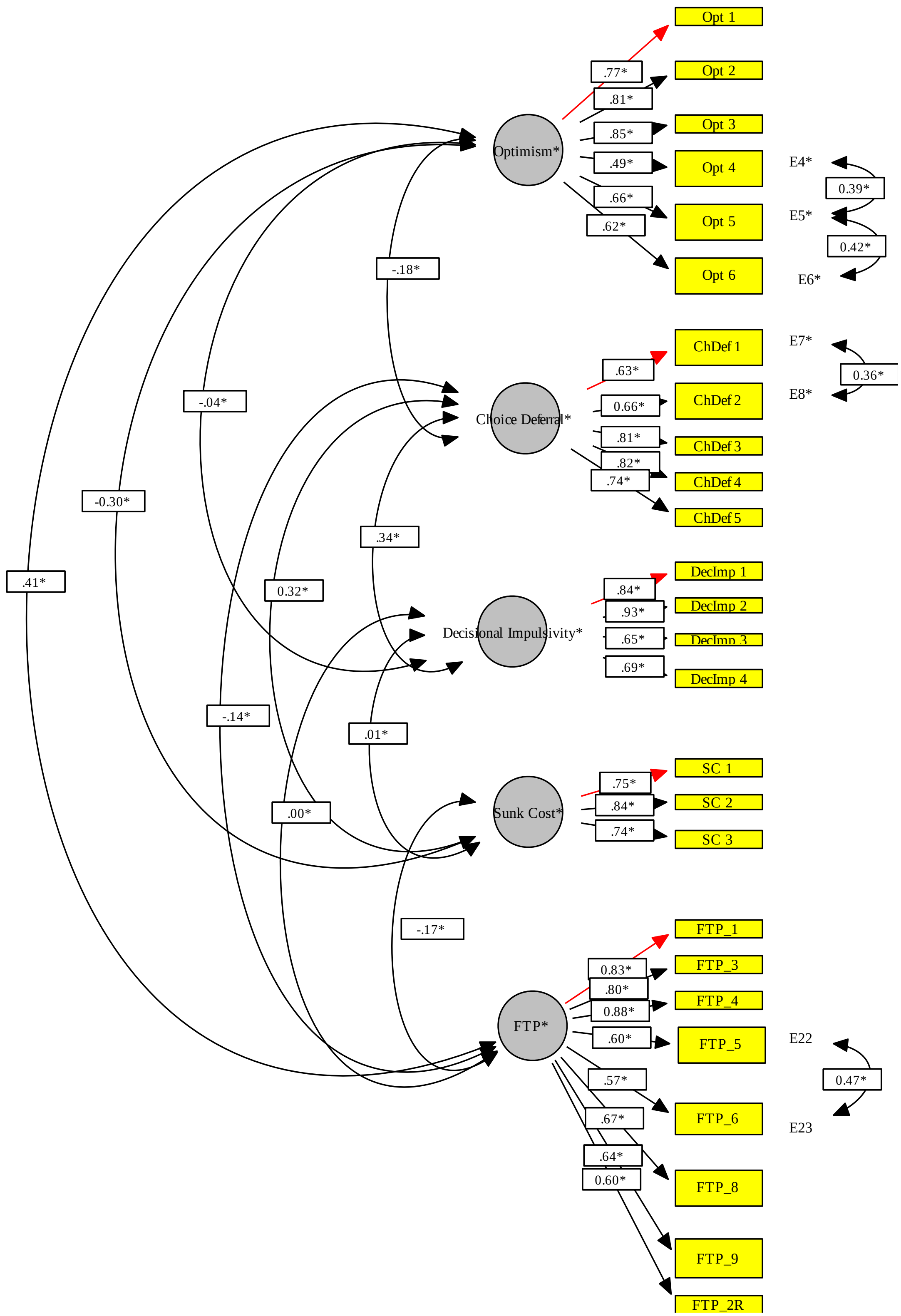

Opt 1
Opt 2
Opt 3
Opt 4
Opt 5
Opt 6
.77*
.81*
.85*
.49*
.66*
.62*
Optimism*
E4*
0.39*
E5*
0.42*
E6*
-.18*
ChDef 1
ChDef 2
ChDef 3
ChDef 4
ChDef 5
E7*
0.36*
E8*
.63*
0.66*
.81*
.82*
.74*
Choice Deferral*
-.04*
-0.30*
.34*
.41*
0.32*
DecImp 1
DecImp 2
DecImp 3
DecImp 4
.84*
.93*
.65*
.69*
Decisional Impulsivity*
-.14*
.01*
.00*
Sunk Cost*
SC 1
SC 2
SC 3
.75*
.84*
.74*
-.17*
FTP_1
FTP_3
FTP_4
FTP_5
FTP_6
FTP_8
FTP_9
FTP_2R
0.83*
.80*
0.88*
.60*
.57*
.67*
.64*
0.60*
FTP*
E22
0.47*
E23

Figure 16-B. Standardized solution for measurement model for base sample (N = 434).

APPENDIX C

MEASUREMENT MODEL TABLES, FIGURES, AND DESCRIPTIVE INTERPRETATIONS FOR CONTRAST SAMPLES

Age – contrast measurement model. The aggregate dataset comprising workers of all ages (*N* =519) was subjected to a highly similar measurement model specification procedure as used in the primary analyses. With the *a-priori* expectation of employing parceling methodology in the auxiliary analyses, however, a more liberal criteria for restricted-backward specification was applied, specifically, standard item loadings below .39 (previously, .34) were excluded. Also, similar to the presentation of primary analyses, tables in the main-text are limited to the listwise sample (*N* = 317), while results for the corresponding DML sample is presented in Appendix C. Results are presented in Table 23 below.

Nested factor-solution comparisons terminated at five factors for comparability to the solutions obtained in the primary analyses. The final measurement model (see, Table 26 below) in the aggregated sample was also examined within each subgroup. Initial inspection indicated a well-fitting model to the observed sample data, with robust estimates of SB-$X^2_{(307)}$ = 483.49, AASR = .05, RMSEA = .05, CFI = .95. The measurement model was also examined in each subgroup, prior to structural model estimation (Bentler, 2006). Results indicated good fit to observed-data in the younger worker sample, with robust estimates of SB-$X^2_{(307)}$ = 483.49, AASR = .05, RMSEA =

.05, CFI = .95, and in the older worker sample, with robust estimates of SB-$X^2_{(307)}$ = 483.49, AASR = .05, RMSEA = .05, CFI = .95.

Table 23

Summary of Restricted Backward-search Measurement Model Respecification for Age-Contrast Analysis in Listwise Sample (N = 317).

Round	**Iter**	**Factor Label**	**# *J***	**$\Lambda_{j,k}$ < .32**	**<Λ_j^2>**	**Item Content**
1	9	Impulsivity	4	$\Lambda_{5,3}$ = .25	.60	When making decisions, I do what seems natural at the moment.
		Sunk Costs	10	$\Lambda_{1,4}$ = .24	-	I would waste time worrying about it.
		Sunk Costs	9	$\Lambda_{3,4}$ = .17	-	It would take me a long time to adjust myself to it.
		Sunk Costs	8	$\Lambda_{4,4}$ = .23	-	I would feel paralyzed.
		Sunk Costs	7	$\Lambda_{5,4}$ = .17	-	I would have trouble doing anything at all.
		Sunk Costs	6	$\Lambda_{6,4}$ = .09	-	I wouldn't know how to deal.
		Sunk Costs	5	$\Lambda_{7,4}$ = .32	.54	I wouldn't have difficulty starting.
		FTP	11	$\Lambda_{10,5}$ = .13	-	I have the sense that time is running out.
		FTP	10	$\Lambda_{11,5}$ = .07	-	As I get older, I begin to experience time as limited.
		FTP	9	$\Lambda_{12,5}$ = .01	-	I feel the importance of time.
		FTP	8	$\Lambda_{7,5}$ = .05	.50	I have limited time left to live my life.
2	6	Sunk Costs	4	$\Lambda_{1,4}$ = .22	-	I would take immediate action to correct it.
		Sunk Costs	3	$\Lambda_{3,4}$ = .20	.59	I would take action rather than just complaining.
3	5	Optimism	5	$\Lambda_{4,1}$ = .37	.56	If something can go wrong for me, it will.
		FTP	7	$\Lambda_{5,5}$ = .34	-	Most of my life lies ahead of me.
		FTP	6	$\Lambda_{6,1}$ =.35	.54	My future seems infinite to me.

Note. Iter = # of Iterations for optimal parameter estimate convergence. # *J* = remaining items following respecification. < Λ_j^2 > = the expected average communality index from remaining indicators.

Table 24

Nested-Measurement Model Comparisons for Age-Contrast Analysis in Listwise Sample (*N* =317)

Model	**S-B X^2**	***Df***	**CFI**	**AASR**	**BIC**	**RMSEA [90% CI]**	**Change Indicators**			
							[1]**ΔS-B X^2 / Δ*Df***	**ΔCFI**	**Δ*BIC***	**Δ*RMSEA***
1-Factor	2212.63	230	.34	.12		.17 [.16 - .17]				
2-Factor	1848.64	229	.46	.14		.15 [.14 - .16]	102.85 / 2*	.12		.02
3-Factor	1192.36	227	.68	.08		.12 [.11 - .12]	125.67 / 3*	.22		.03
4-Factor	715.52	224	.84	.09		.08 [.08 - .09]	178.01* / 4	.16		.04
5-Factor	434.26	220	.93	.05		.06 [.05 - .06]	226.14* / 5	.09		.02

Note. * $p < .00$. 1. Scaled S-B chi-squared differences computed. CI = confidence interval.

Table 26

Summary of Restricted Backward-search Measurement Model Respecification for Job Status-Contrast Analysis in Listwise Sample (N = 486).

Round	**Iter**	**Factor Label**	**# *J***	**$\Lambda_{j,k} < .32$**	**$<\Lambda_j^2>$**	**Item Content**
1	9	Impulsivity	4	$\Lambda_{5,3} = .28$	.60	When making decisions, I do what seems natural at the moment.
		Sunk Costs	10	$\Lambda_{1,4} = .20$	-	I would waste time worrying about it.
		Sunk Costs	4	$\Lambda_{2,4} = .35$	-	I would take immediate action to correct it.
		Sunk Costs	9	$\Lambda_{3,4} = .22$	-	It would take me a long time to adjust myself to it.
		Sunk Costs	8	$\Lambda_{4,4} = .26$	-	I would feel paralyzed.
		Sunk Costs	7	$\Lambda_{5,4} = .16$	-	I would have trouble doing anything at all.
		Sunk Costs	6	$\Lambda_{6,4} = .10$	-	I wouldn't know how to deal.
		Sunk Costs	5	$\Lambda_{7,4} = .32$	-	I wouldn't have difficulty starting.
		Sunk Costs	3	$\Lambda_{8,4} = .35$	.44	I would take action rather than just complaining.
		FTP	11	$\Lambda_{10,5} = .19$	-	I have the sense that time is running out.
		FTP	10	$\Lambda_{11,5} = .11$	-	As I get older, I begin to experience time as limited.
		FTP	9	$\Lambda_{12,5} = .02$	-	I feel the importance of time.
		FTP	8	$\Lambda_{7,5} = .12$	.53	I have limited time left to live my life.

Note. Iter = # of Iterations for optimal parameter estimate convergence. # *J* = remaining items following respecification. $< \Lambda_j^2 >$ = the expected average communality index from remaining indicators.

Job status – contrast- measurement model. The same measurement model specification procedure applied in the age-contrast analysis was adopted for the job status-contrast analysis. Presentation of tables in the main-text are also concordant with the previous contrast-analysis, limited to the listwise sample (N = 486).

Replicating the analytic procedure in the Age-contrast analysis, the nested factor-solution comparisons culminated with five factors for comparability to the solutions obtained in the primary analyses. Also, the final measurement model indicating good fit to the observed-aggregate sample data (SB-X^2 [288, N = 486] = 714.99, $p < .05$, AASR = .04, RMSEA = .06, CFI = .92) was reexamined within each subgroup prior to structural model estimation. Results indicated a well-fitting measurement model to observed data in the worker sample, with robust estimates of SB-X^2 (289, N = 397) = 697.23, $p < .05$, AASR = .04, RMSEA = .06, CFI = .91, and in the non-retiree sample of SB-X^2 (288, N = 89) = 483.49, $p < .05$, AASR = .07, RMSEA = .07, CFI = .90. The slight decrement in fit for the non-retiree sample is within expectation for the comparatively smaller sample size.

Table 27

Nested-Measurement Model Comparisons for Job Status-Contrast Analysis in Listwise Sample (*N* =486)

Model	**S-B X^2**	***Df***	**CFI**	**AASR**	**BIC**	**RMSEA [90% CI]**	**Change Indicators**			
							[1]ΔS-B X^2 / Δ*Df*	**ΔCFI**	**Δ*BIC***	**Δ*RMSEA***
1-Factor	3626.19	299	.39	.11		.15 [.15 - .16]				
2-Factor	2969.53	298	.51	.11		.13 [.13 - .14]	198.59 / 1*	.11		.02
3-Factor	1796.66	296	.73	.07		.10 [.10 - .11]	7798.03 / 2*	.22		.04
4-Factor	1146.81	293	.84	.05		.08 [.08 - .09]	634.63* / 3	.11		.02
5-Factor	757.25	289	.92	.04		.06 [.05 - .06]	304.97* / 4	.08		.02

Note. * *p* < .00. 1. Scaled S-B chi-squared differences computed. CI = confidence interval.

Table 28

Measurement Model Modification Fit Indices and Configural Comparisons for Job Status-Contrast Analysis in Listwise Sample (*N* =486)

Model	**S-B X^2**	***Df***	**CFI**	**AASR**	**BIC**	**RMSEA [90% CI]**	**Change Indicators**			
							[1]LR-X^2	**ΔCFI**	**Δ*BIC***	**Δ*RMSEA***
Base	757.25	289	.92	.04		.06 [.05 - .06]				
1-E Cov	714.99	288	.92	.04		.05 [.05 - .06]	42.26 / 1*	.00		.01
Model	**S-B X^2**	***Df***	**CFI**	**AASR**	**BIC**	**RMSEA [90% CI]**	**[2]ΔS-B X^2/ ΔDf**	**ΔCFI**	**ΔBIC**	**ΔRMSEA**
1-E 1 Fac	3503.03	298	.42	.10		.15 [.15 - .15]	[3]2614.30 / 10*	.50		.09
1-E 4 Fac	1106.30	292	.85	.05		.08 [.07 - .08]	300.61 / 4*	.07		.03

Note. * $p < .00$. 1. LM - Modification Index Computed on Likelihood Ratio Chi-squared Differences. 2. Scaled-Satorra Bentler Chi-Squared Differences Computed. 3. Change Indicators Computed in Reference to Modified Base 5-Factor Model. CI = confidence interval. 2-E = Two errors covaried, 4-E = Four errors covaried. Fac = Factor.

APPENDIX D

SUMMARY ITEM-PARCEL PROCEDURE AND RESULTS

There were many reasons for considering the use of parcels in the current data. First, substantively, the focus of the SEM analyses was on latent structural relations. This is in contrast to the dedicated measurement analysis of the focal construct, FTP (Little, Cuningham, Shahar, & Widaman, 2002). From a model estimation standpoint, parceling was justified by the linear-latent analytic software, again, while not-linearity was inherent to the IRT analysis of the focal FTP construct.

Other advantages to parceling included, for example, distributional affordance for reducing multivariate non-normality as indicated by Mardia’s kappa coefficient (old value, new value – consider table across listwise, DML, cold-deck imputation samples. This is important in making statistical inferences from retro-observation data that may implicate expected findings from the subsequent experimental dataset. That is, it preserves the item-balance of what Little, Lindenberger, & Nesselroade, (1999) term, ‘selection communality’ and ‘selection diversity’, or the measurement error attributable to the instrument and sample, roughly.

Relaxing the assumption of unidimensionality, there is still little basis for conjecturing what and how potential uniqueness or facet-level content might differentially relate to other composite latent factors in the structural model. For this reason, parceling commenced with a domain-representative technique, i.e., distributed uniqueness, and guided by balanced item-factor loading patterns (Hagtvet & Nesser, 2004; Hall, Snell, & Foust, 1999). Specifically, two methods were adopted for forming parcels. First, for latent constructs predicting six or more indicators, weakest loading standard-scored indicators were paired with the strongest loading reverse-scored indicators. This was exhausted for all balance-scored pairs. Second, any items that loaded below the heuristic cut-off for the previous restricted backward-specification procedure was paired with the second-weakest loading indicator, regardless of the number of original indicators predicted by the latent factor. The same minimum of three item- or parcel-indicators in any combination was retained for factor identification in the measurement model.

Table 1-C.

Nested-Measurement –Parcel- Model Comparisons for Listwise Sample (*N* =279)

Model	**S-B X^2**	***Df***	**CFI**	**AASR**	**BIC**	**RMSEA [90% CI]**	**Change Indicators**			
							[1]ΔS-B X^2 / Δ*Df*	**ΔCFI**	**Δ*BIC***	**Δ*RMSEA***
1-Factor	2181.20	252	.33	.13	2316.44	.17 [.16 - .17]				
2-Factor	1715.08	250	.49	.11	1861.49	.15 [.14 - .15]	112.12 / 2*	.02	454.95	.02
3-Factor	1332.34	247	.62	.11	1495.64	.13 [.12- .13]	186.13 / 3*	.00	365.85	.03
4-Factor	1010.79	243	.73	.09	1196.61	.11 [.10 - .11]	171.75* / 4	.02	299.03	.01
5-Factor	457.25	238	.92	.06	671.29	.06 [.05 - .07]	263.74* / 5	.03	525.32	.03
6-Factor	484.26	233	.91	.06	726.39	.06 [.05 - .07]	31.83* / 6	-.01	-55.10	-.01
7-Factor	491.91	225	.91	.07	779.09	.07 [.06 - .07]	8.40 / 7	.00	-52.70	.01

Note. * $p < .00$. 1. Scaled S-B chi-squared differences computed. CI = confidence interval. m = misfit change indicator.

Table 2-C.

Nested-Measurement –Parcel- Model Comparisons for Base Sample (*N* =434)

Model	**S-B X^2**	***Df***	**CFI**	**AASR**	**BIC**	**RMSEA [90% CI]**	**Change Indicators**			
							[1]**ΔS-B X^2 / Δ*Df***	**ΔCFI**	**Δ*BIC***	**Δ*RMSEA***
1-Factor	3477.09	209	.29	.13	3744.21	.19 [.18 - .20]				
2-Factor	2759.89	208	.45	.11	3033.18	.16 [.16 - .17]	678.61 / 1*	.16	711.03	.03
3-Factor	2340.65	206	.54	.12	2626.08	.13 [.12- .13]	155.41 / 2*	.09	407.10	.03
4-Factor	1685.12	203	.68	.09	1988.77	.13 [.12 - .14]	787.88 / 3*	.14	637.31	.00
5-Factor	522.85	199	.93	.04	850.79	.06 [.06 - .07]	2349.99 / 4*	.25	1137.98	.07
6-Factor	514.85	194	.93	.04	873.16	.06 [.06 - .07]	7.76 / 5	.00	-22..37	.00
7-Factor	502.44	188	.93	.04	897.19	.07 [.06 - .07]	11.13 / 6	.00	-24.03	.00

Note. * *p* < .00. 1. Scaled S-B chi-squared differences computed. CI = confidence interval.

Table 9-Par.

Measurement –Parcel- Model Modification Fit Indices and Configural Comparisons for Listwise Sample (N = 279)

Model	**S-B X^2**	***Df***	**CFI**	**AASR**	**BIC**	**RMSEA [90% CI]**	**Change Indicators**			
Base	457.25	238	.92	.06	671.29	.06 [.05 - .07]	**[1]LR-$X^2_{(1)}$**	**ΔCFI**	**Δ*BIC***	**Δ*RMSEA***
2-E Cov	380.14	236	.95	.04	605.38	.05 [.04 - .06]	77.11 / 2*	.03	65.91	.01

							Change Indicators			
Model	**S-B X^2**	***Df***	**CFI**	**AASR**	**BIC**	**RMSEA [90% CI]**	**ΔS-B X² / ΔDf**	**ΔCFI**	**ΔBIC**	**ΔRMSEA**
2-E 1 Fac	1948.22	250	.41	.12	2094.63	.16 [.15 - .16]	1946.41 / 14*	.54	-1489.25	.11
2-E 4 Fac	940.46	241	.76	.09	1137.55	.10 [.10 - .11]	139.53 / 5*	.19	-531.17	.02
2-E 6 Fac	419.66	231	.93	.06	673.06	.05 [.05 - .06]	45.76 / 5*	.02	-67.68	.01

Note. * $p < .00$. 1. Scaled S-B chi-squared differences computed. CI = confidence interval. All change indicators for different factor solutions calculated pairwise with modified Base model. S

Table 9-Par B.

Measurement –Parcel- Model Modification Fit Indices and Configural Comparisons for Base Sample (*N* = 434)

Model	**S-B X^2**	***Df***	**CFI**	**AASR**	**BIC**	**RMSEA [90% CI]**	**Change Indicators**			
							[1]LR-$X^2_{(1)}$	**ΔCFI**	**Δ*BIC***	**Δ*RMSEA***
Base	448.28	216	.86	.04	812.66	.06 [.04 - .06]				
2-E Cov	366.16	214	.91	.04	742.69	.05 [.05 - .06]	82.12 / 2*	.05	69.97	.01
3-E Cov	338.11	213	.93	.04	720.71	.04 [.03 - .04]	28..05 / 1*	.02	21.98	.01

Model	**S-B X^2**	***Df***	**CFI**	**AASR**	**BIC**	**RMSEA [90% CI]**	**Change Indicators**			
							ΔS-B X^2 / ΔDf	**ΔCFI**	**ΔBIC**	**ΔRMSEA**
3-E 1 Fac	1191.11	227	.40	.18	1488.69	.10 [.09 - .10]	1946.41 / 14*	.53	-767.98	.12
3-E 4 Fac	953.04	218	.54	.06	1305.27	.09 [.08 - .09]	614.93 / 5*	.39	-584.56	.02
3-E 6 Fac	344.41	207	.92	.04	763.36	.04 [.03 - .05]	45.76 / 6*	.02	-42.65	.00

Note. Y-B = Yuan-Bentler scaled chi-square, scaled Y-B chi-squared differences computed. CI = confidence interval. All change indicators for different factor solutions calculated pairwise with modified Base model,* p < .00. 1.

Main-Effect Hypothesis Testing

As expected, the parceling technique improved global model fit indices. Initial inspection of the estimated structural model indicated robust estimates of SB-$X^2_{(215)}$ = 353.74, AASR = .05, RMSEA = .05, CFI = .95 to the Listwise-sample data (*N* = 279). The distribution of standardized residuals also appeared approximately symmetric centered about zero. For the DML-estimated sample (*N* = 434), inspection of generalized least squares-based (GLS) tests of homogeneity of means and covariances for the missing-value pattern indicated non-significance, $GLS_{(\bar{x})}$ X^2 (44, 434) = 49.06, $p = .28$, $GLS_{(\Sigma i,j)}$ X^2 (253, 434) = 273.94, $p = .17$. The finding provides support for the missing-value cases and complete cases being sampled from a similarly homogeneous population, also, it permits inferences of reduced nonresponse bias from obtained parameter estimates through DML estimation within a single structural model.

R-H1a**:** Optimism will positively relate to FTP.
R-H1b**:** Optimism will positively relate to intended retirement.

Inspection of the latent path coefficient between latent variables Optimism and FTP indicated a significant positive parameter estimate γ_{51} = .43 (.01), $t(215) = 4.47$, $p < .00$, providing support for hypothesis 1a. Inspection of the structural path coefficient between latent variable Optimism and measured variable Intended Retirement was neither significant nor in the hypothesized direction, providing no support for hypothesis 1b, $\gamma = -.04$ (.15), $t(215) = .15$, $p > .05$.

R-H2a**:** Choice deferral will negatively relate to FTP.
R-H2b**:** Choice deferral will negatively relate to intended retirement.

Inspection of the structural path coefficient between latent variables Choice Deferral and FTP indicated a significant negative parameter estimate γ_{52} = -.69 (.30), $t(215) = -2.28$, $p < .00$, providing support for hypothesis 2a. The value was found to be non-significant in the base-DML sample, γ_{52} = -.13 (.09), $t(215) = -1.32$, $p > .05$.

Inspection of the structural path coefficient between latent variable Choice Deferral and measured variable Intended Retirement indicated a negligible relationship and provided no support for hypothesis 2b, γ = .01 (.58), $t(215) = .02$, $p > .05$.

R-H3a**:** Decisional impulsivity will positively relate to FTP.
R-H3b**:** Decisional impulsivity will positively relate to intended retirement.

Inspection of the structural path coefficient between latent variables Decisional Impulsivity and FTP indicated a significant parameter estimate in the hypothesized direction, providing support for hypothesis 3a, γ_{53} = .48 (.21), $t(215) = 2.27$, $p < .05$. Inspection of the structural path coefficient between latent variable Decisional Impulsivity and measured variable Intended Retirement indicated no relation, γ = .00 (.39), $t(215) = .00$, $p > .05$) and failed to provide support for hypothesis 3b.

R-H4a**:** Sunk costs will negatively relate to FTP.
R-H4b**:** Sunk costs will negatively relate to intended retirement.

Inspection of the structural path coefficient between latent variables Sunk Cost and FTP indicated a weak, non-significant relationship in the hypothesized direction, but provided no support for hypothesis 4a, γ_{54} = -.02 (.20), $t(215) = -.11$, $p > .05$. Inspection of the structural path coefficient between latent variable Sunk Cost and measured variable Intended Retirement was also non-significant, further, counter to the hypothesized direction and provided no support for hypothesis 4b, γ = .11 (.33), $t(215) = .32$, $p > .05$. The obtained parameter trended toward significance in the base-DML sample, γ = .74 (.41), $t(351) = 1.80$, $p = .07$.

Interaction-Effect Hypothesis Testing

Initial inspection of the multigroup structural model with freely estimated factor loadings constrained equivalent across groups indicated a fairly well-fitting measurement model to the observed sample data with robust estimates, SB-$X^2(451, 279) = 567.09$, $p <$ .00, AASR = .06, RMSEA = .05, CFI = .94, BIC = -637.49.

***R-H6a*: **SLE will interact with Optimism in predicting FTP, such that SLE will increase the positive effect.

A backward-search test proceeded with an added equality constraint imposed on the latent path coefficient between Optimism and FTP, and the model was reestimated.

Inspection of the univariate incremental Wald test for dropping the equality constraint of the latent path coefficient between Optimism and FTP indicated significance with LR-$X^2_{(1)} = 32.68$, $p < .00$. On releasing the equality constraint, the multigroup model was reestimated and results indicated comparable statistical fit to the sample data, ΔSB-$X^2_{(1)} = 0.18$, $p = .78$, ΔAASR = .00, ΔRMSEA = .00, ΔCFI = .00, while Bayesian information criterion indicated improved fit ΔBIC = -3.15 Finally, inspection of the relative magnitude of the standardized latent path coefficients indicated that the prediction was stronger in the high-SLE group, β = .48(.15) vs .38(.17), which provided support for hypothesis 6a.

R-H7a**:** SLE will interact with Choice Deferral in predicting FTP, such that SLE will buffer the negative effect.

Inspection of the univariate incremental Wald test for dropping the equality constraint of the latent path coefficient between Choice Deferral and FTP indicated significance with LR-$X^2_{(1)}$ = 3.98, p = .04. On releasing the equality constraint, the multigroup model was reestimated and results indicated comparable statistical fit to the sample data, ΔSB-$X^2_{(1)}$ = .65, p = .42, ΔAASR = .00, ΔRMSEA = .00, ΔCFI = .00, while Bayesian information criterion indicated improved fit ΔBIC = -3.48. Finally, inspection of the relative magnitude of the standardized latent path coefficients indicated that the prediction was weaker in the high-SLE group, γ_{52} = -.12(.20) vs -.17(.20), which provided support for hypothesis 7a.

R-H8a**:** SLE will interact with Decisional Impulsivity in predicting FTP, such that SLE will increase the positive effect.

Inspection of the univariate incremental Wald test for dropping the equality constraint of the latent path coefficient between Decisional Impulsivity and FTP indicated significance with LR-$X^2_{(1)}$ = 3.98, p = .04. On releasing the equality constraint, the multigroup model was reestimated and results indicated comparable statistical fit to the sample data, ΔSB-$X^2_{(1)}$ = 0.08, p = .81, ΔAASR = .00, ΔRMSEA = .00, ΔCFI = .00, while Bayesian information criterion indicated improved fit ΔBIC = -3.48. Finally, inspection of the relative magnitude of the standardized latent path coefficients indicated that the prediction was weaker in the high-SLE group, β = -.12(.20) vs -.17(.20), which provided support for hypothesis 6a.

R-H9a**:** SLE will interact with Sunk Costs in predicting FTP, such that SLE will buffer the negative effect.

Inspection of the univariate incremental Wald test for dropping the equality constraint of the latent path coefficient between Choice Deferral and FTP indicated significance with LR-$X^2_{(1)}$ = 3.98, p = .04. On releasing the equality constraint, the multigroup model was reestimated and results indicated comparable statistical fit to the sample data, ΔSB-$X^2_{(1)}$ = 2.07, p = .15, ΔAASR = .00, ΔRMSEA = .00, ΔCFI = .00, while Bayesian information criterion indicated improved fit ΔBIC = -3.48. Finally, inspection of the relative magnitude of the standardized latent path coefficients indicated that the prediction was weaker in the high-SLE group, β = -.12(.20) vs -.17(.20), which provided support for hypothesis 6a.

APPENDIX E

EXPERIMENT SURVEY STUDY MEASURES

Demographics

1) Please enter your age _____
2) Please select your sex below
   - Male
   - Female

3) Please indicate which of the following best represents your racial or ethnic heritage.
   - Non-Hispanic White or Euro-American
   - Black, Afro-Caribbean, or African American
   - Latino or Hispanic American
   - East Asian or Asian American
   - South Asian or Indian American
   - Middle Eastern or Arab American
   - Native American or Alaskan Native
   - Hawaiian / Pacific Islander

[page break]

Personality

Here are a number of personality traits that may or may not apply to you. Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement. You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other.

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

I see myself as:

1. _____ Extraverted, enthusiastic.

2. _____ Critical, quarrelsome.

3. _____ Dependable, self-disciplined.

4. _____ Anxious, easily upset.

5. _____ Open to new experiences, complex.

6. _____ Reserved, quiet.

7. _____ Sympathetic, warm.

8. _____ Disorganized, careless.

9. _____ Calm, emotionally stable.

10. _____ Conventional, uncreative.

[page break]

Personality Vignettes

Next, you will see a passage that describes a type of person. After reading the passage, please respond to the items below it, rating the person you read about on the following 7-point rating scale:

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

Thomas/Dorothy takes charge and is the first to act. He/she knows how to captivate people and can express his/her self easily. He/she possesses leader-like qualities. He/she loves large parties. He/she starts conversations easily and ends up talking to a lot of different people at parties. He/she is typically the life of the party.

I see Thomas / Dorothy as

1. _____ Extraverted, enthusiastic.
2. _____ Critical, quarrelsome.
3. _____ Dependable, self-disciplined.
4. _____ Anxious, easily upset.
5. _____ Open to new experiences, complex.
6. _____ Reserved, quiet.
7. _____ Sympathetic, warm.
8. _____ Disorganized, careless.
9. _____ Calm, emotionally stable.
10. _____ Conventional, uncreative.

Next, you will see another passage that describes another type of person. After reading the passage, please respond to the item below it, rating the person you read about on the following 7-point rating scale:

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

James/Mary doesn't talk a lot. He/she doesn’t like to draw attention to him/herself so he/she keeps in the background and remains quiet around strangers. He/she tends to keep his/her thoughts and feelings to his/her self. He/she is the type of person who lets others make the decisions and lead the way. He/she has difficulty expressing his/her feelings and is afraid to draw attention to his/her self.

I see James / Mary as

1. _____ Extraverted, enthusiastic.
2. _____ Critical, quarrelsome.
3. _____ Dependable, self-disciplined.
4. _____ Anxious, easily upset.
5. _____ Open to new experiences, complex.
6. _____ Reserved, quiet.
7. _____ Sympathetic, warm.
8. _____ Disorganized, careless.
9. _____ Calm, emotionally stable.
10. _____ Conventional, uncreative.

Next, you will see another passage that describes another type of person. After reading the passage, please respond to the item below it, rating the person you read about on the following 7-point rating scale:

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

Robert/Jennifer doesn't plan ahead and tends to leave his/her work undone. He/she often takes things too lightly and ends up neglecting his/her responsibilities. He/she tends to make rash decisions, often doing the opposite of what is asked. He/she disregard rules and occasionally does improper things.

I see Robert / Jennifer as

1. _____ Extraverted, enthusiastic.
2. _____ Critical, quarrelsome.
3. _____ Dependable, self-disciplined.
4. _____ Anxious, easily upset.
5. _____ Open to new experiences, complex.
6. _____ Reserved, quiet.
7. _____ Sympathetic, warm.
8. _____ Disorganized, careless.
9. _____ Calm, emotionally stable.
10. _____ Conventional, uncreative.

Next, you will see another passage that describes another type of person. After reading the passage, please respond to the item below it, rating the person you read about on the following 7-point rating scale:

_

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

Joseph/Margaret behaves properly, following directions and sticking to the rules. He/she also appreciate good manners. He/she completes tasks on time and according to plan. He/she usually prepares for things ahead of time to avoid making mistakes. He/she actually likes to plan ahead.

I see

1. _____ Extraverted, enthusiastic.
2. _____ Critical, quarrelsome.
3. _____ Dependable, self-disciplined.
4. _____ Anxious, easily upset.
5. _____ Open to new experiences, complex.
6. _____ Reserved, quiet.
7. _____ Sympathetic, warm.
8. _____ Disorganized, careless.
9. _____ Calm, emotionally stable.
10. _____ Conventional, uncreative.

Next, you will see another passage that describes another type of person. After reading the passage, please respond to the item below it, rating the person you read about on the following 7-point rating scale:

_

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

David/Barbara has no problems working with complex problems or comprehending a lot of information. He/she usually gets the idea of things right away. He/she typically catches on to things quickly. He/she enjoy learning things and finds abstract ideas interesting.

He/she also have a rich vocabulary and shows a mastery of language.

I see David / Barbara as

1. _____ Extraverted, enthusiastic.
2. _____ Critical, quarrelsome.
3. _____ Dependable, self-disciplined.
4. _____ Anxious, easily upset.
5. _____ Open to new experiences, complex.
6. _____ Reserved, quiet.
7. _____ Sympathetic, warm.
8. _____ Disorganized, careless.
9. _____ Calm, emotionally stable.
10. _____ Conventional, uncreative.

Next, you will see another passage that describes another type of person. After reading the passage, please respond to the item below it, rating the person you read about on the following 7-point rating scale:

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

William/Linda is not interested in abstract ideas. He/she generally dislikes learning and has a poor vocabulary. He/she has a hard time understanding complex things so he/she avoids them. For example, he/she tends to put off reading material that is complex and steers clear of complicated people.

I see William / Linda as

1. _____ Extraverted, enthusiastic.
2. _____ Critical, quarrelsome.
3. _____ Dependable, self-disciplined.
4. _____ Anxious, easily upset.
5. _____ Open to new experiences, complex.
6. _____ Reserved, quiet.
7. _____ Sympathetic, warm.
8. _____ Disorganized, careless.
9. _____ Calm, emotionally stable.
10. _____ Conventional, uncreative.

Next, you will see another of passage that describe another type of person. After reading the passage, please respond to the item below it, rating the person you read about on the following 7-point rating scale:

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

Charles/Jessica avoids conflicts and criticizing other people. He/she respects and sympathizes with others' feelings and appreciates their viewpoints.

I see Charles / Jessica as

1. _____ Extraverted, enthusiastic.
2. _____ Critical, quarrelsome.
3. _____ Dependable, self-disciplined.
4. _____ Anxious, easily upset.
5. _____ Open to new experiences, complex.
6. _____ Reserved, quiet.
7. _____ Sympathetic, warm.
8. _____ Disorganized, careless.
9. _____ Calm, emotionally stable.
10. _____ Conventional, uncreative.

Next, you will see another passages that describes another type of person. After reading the passage, please respond to the item below it, rating the person you read about on the following 7-point rating scale:

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

John/Patricia takes no time for others and is not interested in other people's problems. He/she is generally indifferent to the feelings of others. He/she is not afraid of conflict or confrontation. He/she values competition over cooperation.

I see John / Patricia as

1. _____ Extraverted, enthusiastic.
2. _____ Critical, quarrelsome.
3. _____ Dependable, self-disciplined.
4. _____ Anxious, easily upset.
5. _____ Open to new experiences, complex.
6. _____ Reserved, quiet.
7. _____ Sympathetic, warm.
8. _____ Disorganized, careless.
9. _____ Calm, emotionally stable.
10. _____ Conventional, uncreative.

Next, you will see another passage that describes another type of person. After reading the passage, please respond to the item below it, rating the person you read about on the following 7-point rating scale:

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

Michael/Elizabeth gets upset and stressed out easily. His/her mood changes a lot. He/she often feels blue and is easily threatened. He/she is filled with doubts about things and dislikes his/her self.

I see Michael / Elizabeth as

1. ______ Extraverted, enthusiastic.
2. ______ Critical, quarrelsome.
3. ______ Dependable, self-disciplined.
4. ______ Anxious, easily upset.
5. ______ Open to new experiences, complex.
6. ______ Reserved, quiet.
7. ______ Sympathetic, warm.
8. ______ Disorganized, careless.
9. ______ Calm, emotionally stable.
10. ______ Conventional, uncreative.

Next, you will see another passages that describes another type of person. After reading the passage, please respond to the item below it, rating the person you read about on the following 7-point rating scale:

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

Richard/Susan is the type of person who looks on the bright side of life. He/she feels comfortable with his/her self. He/she is not easily frustrated or bothered by things, and he/she seldom takes offense.

I see Richard / Susan as

1. ______ Extraverted, enthusiastic.
2. ______ Critical, quarrelsome.
3. ______ Dependable, self-disciplined.
4. ______ Anxious, easily upset.
5. ______ Open to new experiences, complex.
6. ______ Reserved, quiet.
7. ______ Sympathetic, warm.
8. ______ Disorganized, careless.
9. ______ Calm, emotionally stable.
10. ______ Conventional, uncreative.

<page break>

Next, you will see another passage that describes another type of person. After reading the passage, please respond to the item below it, using the following 7-point rating scale:

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

GREGARIOUSNESS. Gregariousness reflects the degree to which a person tends to socialize with others. Generally, people who describe themselves as gregarious attend parties and social gatherings, and feel comfortable starting conversations with just about anyone.

1) I see myself as a gregarious person. ___

<page break>

Next, you will see another passage that describes another type of person. After reading the passage, please respond to the item below it, using the following 7-point rating scale:

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

CONSCIENTIOUSNESS. Conscientiousness reflects a tendency to do what is considered right and complete tasks thoroughly. Conscientious people say they generally do things according to a plan, and try to avoid making mistakes and leaving their work unfinished.

1) I see myself as a conscientious person. ___

<page break>

Next, you will see another different passage that describes another type of person. After reading the passage, please respond to the item below it, using the following 7-point rating scale:

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

INTELLECTUAL. Intellectual refers to the capacity to understand abstract concepts. People who describe themselves as intellectual say they have rich vocabularies, love learning about new things, and can make insightful remarks.

1) I see myself as an intellectual person. ___

<page break>

Next, you will see another passage that describes another type of person. After reading the passage, please respond to the item below it, using the following 7-point rating scale:

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

UNDERSTANDING. Understanding reflects the degree to which a person has a sympathetic awareness for other people. Understanding individuals describe themselves as respectful, helpful, and sympathetic towards others' feelings.

1) I see myself as an understanding person. ___

<page break>

Next, you will see another passage that describes another type of person. After reading the passage, please respond to the item below it, using the following 7-point rating scale:

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

STABILITY. Stability reflects the extent to which an individual is emotionally steady. People who are stable tend to keep their cool, seldom take offense, or become bothered by things.

1) I see myself as a stable person. ___

Retirement Planning

Below is a list of retirement preparation activities. Please read carefully each behavior statement and indicate, using the 5-point scale, below, please indicate the extent to which you engage in each activity.

Never	Rarely	Sometimes	Fairly Often	Almost always
1	2	3	4	5

1. Seeking financial advice about retirement issues.

2. Reading articles or attending seminars about planning retirement.

3. Keeping up with tax and other financial changes that might affect your retirement.

4. Topping up your savings for retirement *(beyond required superannuation).*

5. Adjusting your spending habits to prepare for retirement.

6. Working on developing a circle of friends for your retirement years.

7. Developing interests, hobbies and skills now that you can use in retirement.

8. Developing plans for what you might do in retirement.

9. Thinking about what might be the personal challenges for you in your retirement.

10. Developing 'back up' plans in the case the unexpected happens *(e.g. health problems, loss of job).*

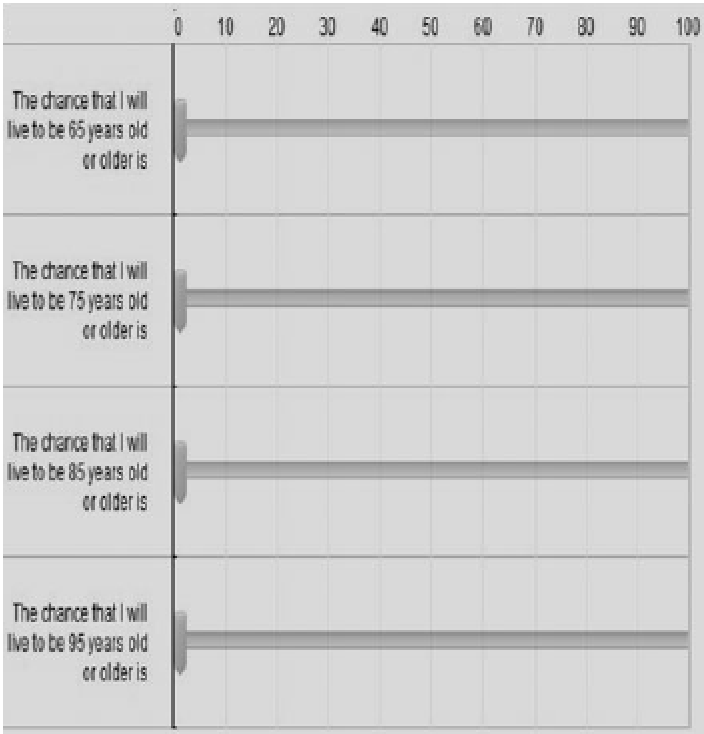
11. Changing your expectations to match what you think the opportunities will be in your retirement.

[page break]

Subjective Life Expectancy

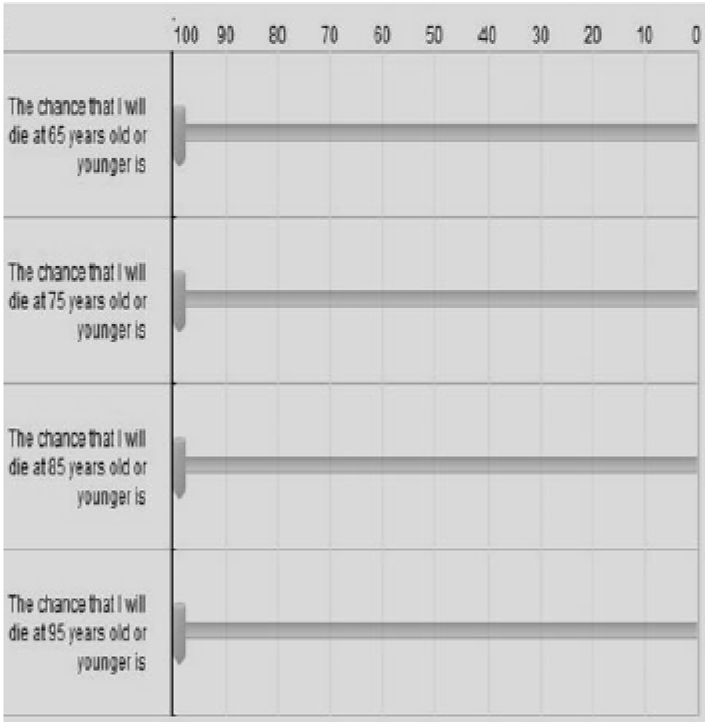
<u>'Live-to' condition.</u>

Using the slider scale below, please indicate the approximate age you currently believe that you will live to.



<u>'Die-by' condition.</u>

Using the slider scale below, please indicate the approximate age you currently believe that you will die by.



[page break]

Future Time Perspective

Read each item and, as honestly as you can, answer the question: "How true is this of you?"
Shade the oval under the appropriate number on the scale, where 1 means the statement is
very untrue for you and 7 means that the statement is very true for you.

Very untrue of me	Untrue of me	Somewhat untrue of me	Neither true or untrue of me	Somewhat true of me	True of me	Very true of me
1	2	3	4	5	6	7

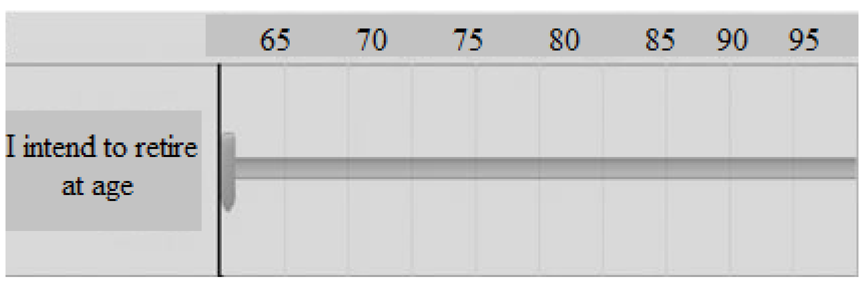
1) Many opportunities await me in the future.

2) I expect that I will set many new goals in the future.
3) My future is filled with possibilities.
4) Most of my life lies ahead of me.
5) My future seems infinite to me.
6) I could do anything I want in the future.
7) There is plenty of time left in my life to make new plans.
8) I have the sense that time is running out.
9) There are only limited possibilities in my future.
10) As I get older, I begin to experience time as limited.
11) I have limited time left to live my life.
12) I feel the importance of time.
13) I enjoy thinking about how I will live years from now in the future.
14) The distant future is too uncertain to plan for. (r)
15) The future seems very vague and uncertain to me. (r)
16) I pretty much live on a day-to-day basis. (r)
17) I enjoy living for the moment and not knowing what tomorrow will bring. (r)

[page break]

Retirement Intentions

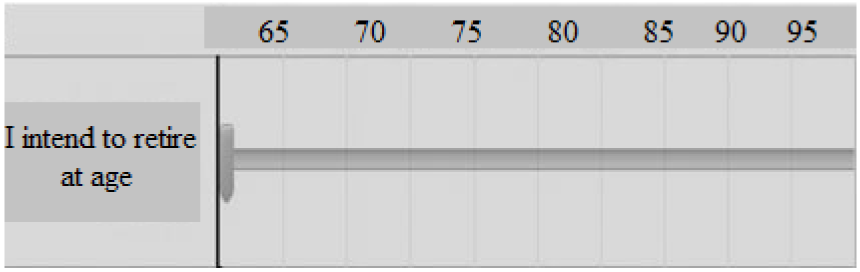
1) *Now, using the slider scale below, please indicate the age you intend to retire.*



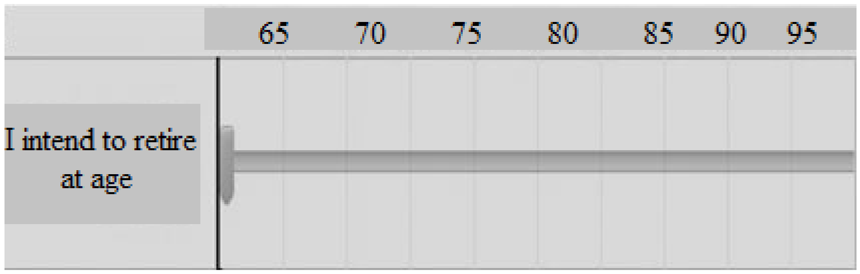
[page break]

Now, using the slider scales below, please indicate the age you intend to retire from each scenario. If the scenario does not distinguish from your general retirement plan, please skip it and move on to the next scenario.

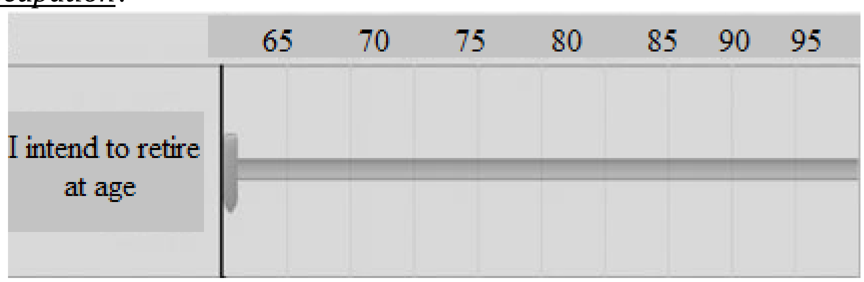
2) *Using the slider scale below, please indicate the age you intend to retire from your job or employer?*



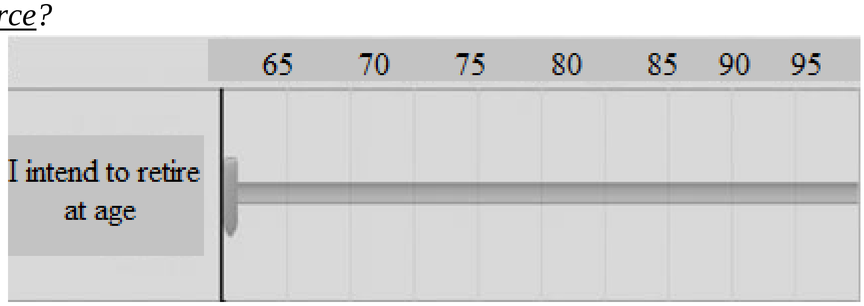
3) *Using the slider scale below, please indicate the age you intend to retire from your <u>career</u>?*



4) *Using the slider scale below, please indicate the age you intend to retire from your <u>occupation</u>?*



5) *Using the slider scale below, please indicate the age you intend to retire from the <u>workforce</u>?*



[page break]

Your survey session is complete. Thank you for your participation.

APPENDIX F

CURRICULUM VITAE

Matthew J. Kerry
Doctoral Candidate
School of Psychology | Georgia Institute of Technology
654 Cherry St NW | Atlanta, GA 30332-0170
Phone: 678.787.1838 | E-Mail: mkkerry@gmail.com

Education

Emory University, Atlanta, GA

2008 B.A. Psychology

B.A. Economics

Georgia Institute of Technology, Atlanta, GA

2012 M.S. Industrial / Organizational Psychology

Thesis – "Person and Professional Program Determinants of Health Provider Student Attitudes toward Inter-Professional Teamwork"

2014 Ph.D. Industrial / Organizational Psychology, *expected* 12/14

Thesis – „Construct Repre
Minor: Quantitative Psychology

Federal Technical College (ETH), Zurich, Switzerland

2014 Post-doc, *selected – acceptance TBD*

Research Interests

Teamwork determinants, dynamics, and performance; Multi-team system effectiveness

Risk and uncertainty in high-stakes industry, in teams, and over time

Health education and training assessment; Care delivery innovation

Organizational culture change management

Peer-Reviewed Publications

Kerry, M. J., Posnock, S., Kanfer, R., & Ander, D. (2014; *forthcoming submission*). Team- process and –outcome efficacy agreement over time. *Organizational Research Methods*.

Kerry, M. J. & Kanfer, R., & Ander, D. (2014; *forthcoming submission*). A latent variable approach to pre-licensure professional identity. *Journal of Interprofessional Care*.

Kerry, M. J., Posnock, S, Ander D., & Robertson, B. (2014; *forthcoming submission*). The primrose path of program differences in interprofessional education: Empirical support of relevant student differences and future considerations. *Journal of Interprofessional Care.*

Book Chapters and Published Reviews

Kerry, M. (2013). Motivation in Human-Systems Integration [Review of the chapter *Motivation*, by J. B. Vancouver], in In D.Boehm-Davis, J. D. Lee, & F. T. Durso (Eds). *Handbook of Human-Systems Integration*.

Kanfer, R. & **Kerry, M.** (2011). Motivation in multiteam systems. In S.J. Zaccaro, M.A. Marks, & L. DeChurch (Eds.), *Multiteam Systems: An Organization Form for Dynamic andComplex Environments:* 81-108. New York: Taylor & Francis.

Conference Proceedings

Kerry, M. J., (2014). *Conceptualization and development of a measure of patient identity*. Poster accepted for presentation at the 9th annual meeting of the Interdisciplinary Network for Group Research, Raleigh, NC.

DiazGrandos, D. & **Kerry, M.** (co-chairs) (2014). *I/O Psychology Perspectives on Advances in the Healthcare Industry*. Symposium to-be submitted for presentation at the 74th annual meeting for the Academy of Management, Philadelphia, PA.

Kerry, M. & Lazzara, E. (co-chairs) (2014). *A Look at a Paradigmatic Shift in Healthcare Training*. Symposium accepted for presentation at the 29th annual meeting for the Society for Industrial / Organizational Psychology, Honolulu, HI.

Kerry, M. (2013, February). Student and Program Influences on Interprofessional Education Outcomes. To-be presented at the *Georgia Tech Research and Innovation Conference*, Atlanta, GA.

Kerry, M., Dainis, A., & Kantrowitz, T. (2012, April). Cross-cultural Biodata: Toward a Common Ground. Presented at the 27th Annual Meeting of the *Society for Industrial & Organizational Psychology,* San Diego, CA.

Kanfer, R., **Kerry, M.**, Posnock, S., Robertson, B., & Ander, D. S. (2011, July). Interprofessional Team Training: Person and Disciplinary Determinants of Training Outcomes. Presented at the 15th Biennial Meeting of the *International Society for the Study of Individual Differences*, London, UK.

Academic, Research, and Teaching Experience Summer 2013

Instructor Adult Abnormal Psychology

Fall 2009 – Current

Knowledge and Skills Lab, School of Psychology, Georgia Institute of Technology

Guest Lecturer

- Intro psychology
- I/O psychology

Graduate Research Assistant

- Inter-professional training design, delivery, and assessment
- Test construction
- Data collection and analyses, technical writing

Graduate Teaching Assistant

- Introduction to Psychology
- Personality Psychology
- Industrial and Organizational Psychology

Summer 2008

Neuro-Behavioral Economics Lab, Department of Economics, Emory University

Research Assistant

- Literature review and instrument construction for study examining individual predictors of economic and non-economic risky decision-making
- Conduct experimental sessions

Employment

March, 2013 – Current **SimpleC, LLC**

Research Coordinator

- NIH grant implementation
- Study manual technical writing
- Data management and analyses

May, 2012 – March, 2013 **WellStar Health System, Inc.**

Organizational Learning – External Consultant

- Observational measure design, pilot, and implementation
- Structured interviews
- Data management

July 2011 – December 2011 **SHL, Inc.**

Associate Research Scientist (Intern)

- Product Development & Innovation Team
- Assessment validation
- Technical writing

Sept 2007-December 2008 **Knowledge and Skills Lab**, Department of Psychology, Georgia Institute of Technology

Undergraduate Research Lab Coordinator

- Data entry
- Generic item creation
- Participant recruitment

May 2006-June 2007 **Emory Autism Center**, Department of Psychiatry, Emory University

Preschool Teacher

- Integration-education and socializing for children with autism-spectrum disorders
- Linguistic development, language approximations, peer-incidental teaching, ecological sampling teaching, in-home development

Technical Competencies

Analytic Software: R, SPSS, EQS, IRTPro, SAS (limited)

Analytic Frameworks: Uni-Multivariate ANOVA, OLS Regression, HLM, E/CFA, SEM, IRT

Consumer Software: MS Suite, Adobe CS

Programming Language: H/XTML, SQL, Python (limited)

Professional Services

American Psychology Association – *Ad-hoc reviewer*

Journal of Interprofessional Care – *Ad-hoc reviewer*

Medical Education - *Ad-hoc reviewer*

Organizational Behavior and Human Decision Processes – *Ad-hoc Reviewer*

Professional Memberships

American Psychology Association - *Associate member*

Society for Industrial / Organizational Psychology – *Associate member*

Society for Human Factors and Engineering Psychology – *Student affiliate*

Society for the Study of Motivation - *Member*

REFERENCES

Abraham, J. D., & Hansson, R. O. (1995). Successful aging at work: An applied study of selection, optimization, and compensation through impression management. *Journal of Gerontology: Psychological Sciences*, 50B (2), 94 – 103.

Abraham, K. G., & Houseman, S. N. (2004). Work and retirement plans among older Americans (No. 04-105). Upjohn Institute Working Paper.

Ackerman, T. A. (1992). A didactic explanation on item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement*, *29* (1), 67 – 91.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B.N. Petrov & F. Csaki (Eds.), *Second International Symposium on Information Theory* (pp. 267 – 281). Budapest: Akademiai Kiado.

Alexander, C. S., & Becker, H. J. (1978). The use of vignettes in survey research. *Public Opinion quarterly*, *42*(1), 93-104.

American Educational Research Association., American Psychological Association., & National Council on Measurement in Education., & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (1999). *Standards for educational and psychological testing.*

American Educational Research Association., American Psychological Association., National Council on Measurement in Education., & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*.

Anastasi, A. & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, N.J.: Prentice Hall.

Anguinis, H, & Edwards, J. R. (2014). Methodological wishes for the next decade and how to make wishes come true. *Journal of Management Studies, 51*(1), 143-174.

Ashforth, B.E., Harrison, S.H., & Sluss, D.M. (2014). Becoming: The interaction of socialization and identity in organizations over time. In A.J. Shipp & Y. Fried (Eds.), Time and work, vol. 1: How time impacts individuals Philadelphia: Psychology Press.

Bagozzi, R. P. (1984). A prospectus for theory construction in marketing. J. Mark. 48:11–29.

Bagozzi, R. P. & Yi, Y. (1989). On the use of structural equation models in experimental designs. *Journal of Marketing Research*, 271-284.

Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing construct validity in organizational research. *Administrative science quarterly*, 421-458.

Bagozzi, R. P., Yi, Y., & Singh, S. (1991). On the use of structural equation models in experimental designs: Two extensions. *International Journal of Research in Marketing*, *8*(2), 125-140.

Bardo, J. W. & Yeager, S. J. (1982). Consistency of response style across types of Response formats. *Perceptual and Motor Skills*, *55*, 307 - 310.

Bechtold, H. (1959). Construct validity: A critique. *American Psychologist*, *14*, 619 – 629.

Beegle, K., Himelein, K., & Ravallion, M. (2009). Frame-of-reference bias in subjective welfare regressions.

Bennett, R. E. (1990). Toward intelligent assessment: An integration of constructed Response testing, artificial intelligence, and model-based measurement. *Educational Testing Service, Research Report, 90*(5), 1 – 45.

Bentler, P. M. (1990). Comparative fix indexes in structural equation models. *Psychological Bulletin, 107* (2), 238 – 246.

Bentler, P. M. (2006). EQS 6 Structural Equations Program Manual. Encino, CA: Multivariate Software, Inc.

Berg, I. A. (1955). Response bias and personality: The deviation hypothesis. *Journal of Psychology, XL,* 60 – 71.

Bidwell, J., Griffin, B., & Hesketh, B. (2006). Timing of retirement: Including a delay discounting perspective in retirement models. *Journal of Vocational Behavior, 68,* 368-387.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In Lord, F., M., & Novick, M. R. (Eds.), *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.

Bluck, S., & Alea, N. (2009). Thinking and talking about the past: Why remember? *Applied Cognitive Psychology*, *23*(8), 1089-1104.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. C. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*, 700–765.

Bolt, D. & Stout, W. (2007). Differential item functioning: Its multidimensional model And resulting SIBTEST detection procedure. Unpublished manuscript.

Brossman, B. G. (2010). Observed score and true score equating procedures for multidimensional item response theory. Unpublished doctoral dissertation. University of Iowa.

Buss, D. M. (1989). Personality as traits. *American Psychologist, 44*, 1378–1388.

Campbell, D. T. (1960). Recommendations for APA test standards regarding construct, trait or discriminant validity. *American Psychologist, 15,* 546-553.

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56* (2), 81 – 105.

Campbell, D. T., Stanley, J. C., & Gage, N. L. (1963). *Experimental and quasi -experimental designs for research* (pp. 171-246). Boston: Houghton Mifflin.

Carstensen, L. L. and Fredrickson, B. F. Aging, illness and social preferences. Paper presented at the International Congress of Psychology, (1992), Brussels, Belgium.

Carstensen, L. L., & Lang, F. R. (1996). *Future Orientation Scale.* Unpublished manuscript, Stanford University.

Carstensen, L. L. (2006). The influence of a sense of time on human development. *Science*, *312*, 1913-1915.

Casio, W. F. (1995). Whither industrial and organizational psychology in a changing world of work. *American Psychologist*, *50,* 928 - 939.

Cate, R. A., & John, O. P. (2007). Testing models of the structure and development of future time perspective: maintaining a focus on opportunities in middle age. *Psychology and Aging*, *22*(1), 186.

Cattell, R. B. (1957). *Personality and motivation structure and measurement.* New York: World Book, 1957.

Cattell, R. B. (1964). Validity and reliability: A proposed more basic set of concepts.

Journal of Educational Psychology, 55, 1 - 22.

Center for Disease Control and Prevention, (2011). Healthy places terminology, retrieved From http://www.cdc.gov/healthyplaces/terminology.htm on May 5, 2014.

Cermak, L. & Craik, F. (1979). *Levels of Processing in Human Memory*. Hillsdale, NJ: Erlbaum.

Coan, R. W. (1964). Facts, factors, and artifact, the: The quest for psychological meaning, *Psychological Review, 71* (2), 123 – 140.

Cronbach, L. J. (1946). Response sets and test validity. *Educational and psychological measurement*, *6*(4), 475-494.

Cronbach, L. J. (1948). A validation design for qualitative studies of personality. *Journal of Consulting Psychology, 12,* 365 – 374.

Cronbach, L. J. (1950). Further evidence on response sets and test designs. *Education and Psychological Measurement, 10*(3), p. 3 – 31.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443-507). Washington, DC: American Council on Education.

Cronbach, L.J. (1988). Five perspectives on validity argument. In H. Wainer & H. Braun (Eds.), Test validity (pp. 3-17), Hillsdale, NJ: Erlbaum.

Cronbach, L. J. (1989). Construct validation after 30 years. In R. L. Linn (Eds.), *Intelligence:Measurement theory and public policy*). Proceedings of a symposium in honor of Loyd G. Humphreys, pp. 147 – 171. Urbana: University of Illinois Press.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests.

Psychological Bulletin, 52, 281-302.

Dawes, R. M. (1995). Standards of practice. In Hayes, S. C., Follette, V. M., Dawes, R. M., and Grady, K. E. (Eds). *Scientific Standards of Psychological Practice: Issues And Recommendations*. Reno, Nevada: Context Press.

de Groot, W. (2000). Adaptation and scale of reference bias in self-assessments of quality of life. *Journal of health economics*, *19*(3), 403-420.

Denton F. T. & Spencer B. G. (2009) What is retirement? A review and assessment of Alternative concepts and measures. *Canadian Journal of Aging* 28(1): 63–76.

De Boeck, P. & Wilson, M. (Eds.). (2004). *Explanatory item response models: A Generalized linear and nonlinear approach*. New York: Springer.

DeMeyer, I. & Raedt, R. D. (2013). The effect of future time perspective manipulation on Affect and attentional bias. *Cognitive Therapy Research, 38*(3), 302-312.

Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in An ecological approach to cognition. *Psychological Bulletin, 130*, 959 - 988.

Dimitrov, D. M. (2006). Comparing groups on latent variables: A structural equation Modeling Approach. *Work: A Journal of Prevention, Assessment and Rehabilitation*, 26(4), 429-436.

Donahue, W., Orbach, H. L., & Pollak, O. (1960). Retirement: The emerging social pattern. In Tibbits, C. (Ed.), *Handbook of Social Gerontology*, Chicago: University of Chicago Press, 330-406.

Dunnette, M.D. (1966). Fads, fashions and folderol in psychology. *American Psychologist, 12*, 343-352

Edwards J. R. (2008). To prosper, organizational psychology should overcome methodological barriers to progress. *Journal of Organizational Behavior, 29,* 469 – 491.

Edwards, J. R., & Berry, J. W. (2010). The presence of something or the absence of nothing: Increasing theoretical precision in management research. *Organizational Research Methods*, *13*(4), 668-689.

Ekerdt, D. J. (2004). Born to retire: The foreshortened life course. *The Gerontologist, 44* (1), 3 –9.

Ekerdt, D. J. (2010). Frontiers of research on work and retirement. *Journal of Gerontology: Social Sciences*, *65B* (1), 69 – 80.

Ekerdt, D. J., DeViney, S., & Kosloski, K. (1996). Profiling plans for retirement. *Journal Of Gerontology: Social Sciences*, *51B* (3), S140 – S149.

Ekerdt, D. J., Hackney, J., Kosloski, K., & DeViney, S. (2001). Eddies in the Stream The Prevalence of Uncertain Plans for Retirement. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *56*(3), S162-S170.

Embretson, S. E. (Whitely). (1983). Construct validity: Construct rep-resentation versus nomothetic span. Psychological Bulletin, 93(1), 179-197.

Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. *Psychological Methods*, *3*(3), 380.

Embretson, (1999). Generating items during testing: Psychometric issues and models. *Psychometrika, (64)*, 407 – 434.

Embretson, S. E. & Poggio, J. (2012). The impact of scaling and measurement methods On individual differences in growth. In L. Laursen, T. Little n& N. Card (Eds). *Handbook of Developmental Research Methods*. New York: Guilford Press.

Embretson, S. E. & Reise, S. (2000). Item response theory for psychologists. Mahwah, NJ: Erlbaum Publishers.

Embretson, S. E. (2007). Construct validity: A universal validity system of just another Test evaluation procedure? *Educational Researcher, 36* (8), 449 – 455.

Feifel, H. (1957). Judgment of time in younger and older persons. *Journal of Gerontology*, *12*, 71-74.

Feldman, D. C. & Beehr, T. A. (2011). A three-phase model of retirement decision making. *American Psychologist, 66* (3), 193 – 203.

Ferster, A. E. (2013). An evaluation of item level cognitive supports via a random-effects extension of the Linear Logistic Test Model. (Dissertation, University of Georgia). Retrieved from http://athenaeum.libs.uga.edu/handle/10724/29032.

Filer, R. K. and Petri, P. A. (1988). A job-characteristics theory of retirement. *The Review Of Economic and Statistics, 70* (1), 123 - 128.

Finch, H. (2010). Item Parameter Estimation for the MIRT Model Bias and Precision of Confirmatory Factor Analysis—Based Models. *Applied Psychological Measurement, 34*(1), 10-26.

Fiske, D. W. (1982). Convergent-discriminant validation in measurements and research strategies. In D. Brinberg & L. Kidder (Eds.), *New directions for methodology of social and behavioral science: Forms of validity in research.* 12, pp. 79 – 92. San Francisco: Jossey-Bass.

Frank, L. K. (1939). Time perspectives. *Journal of Social Philosophy, 4,* 293 - 312.

Fraser, J. T. The voices of time: A cooperative survey of man's views of time as expressed by the sciences and the humanities, New York: Braziller, 1966 and

London.

Fraser, J. T. The interdisciplinary study of time. *Annals of the New York Academy of Sciences, 138,* 822 – 847.

Fraser, J. T., Haber, F. C., & Müller, G. H. *The Study of Time I*, New York: Springer Verlag, 1972.

Fraser, J. T. & Lawrence, N. *The Study of Time I*, New York: Springer Verlag, 1975.

Fraser, J. T., Lawrence, N., & Park, D. *The Study of Time I,* New York: Springer Verlag, 1978.

Frederiksen, N. Mislevy, R. J., & Bejar, I. I. (Eds.) *Test theory for a new generation of tests.* (1993). Hillsdale, NJ: Erlbaum.

Friedman, E. A., & Havinghurst, R. J. The meaning of work and retirement. Chicago: University of Chicago Press, 1954.

Friedman, S. L. & Scholnick, E. K. (Eds.) The Developmental Psychology of Planning. NY: Psychology Press, 1997.

Goldberg, L. R. & Slovic, P. (1967). Importance of test item content: An analysis of a Corollary of the deviation hypothesis. *Journal of Counseling Psychology*, *14,* 462 – 472.

Granrose, C. S., & Baccili, P. A. (2006). Do psychological contracts include boundaryless or protean careers? *Career Development International*, *11*, 163 – 182.

Griffin, B., & Hesketh, B. (2008). Postretirement work: A model of the determinants of paid and volunteer work. *Journal of Occupational and Organizational Psychology*, *81*(1), p.101-121.

Griffin, B., Hesketh, B., & Loh V., (2012). The influence of subjective life expectancy on retirement transition and planning: A longitudinal study. *Journal of Vocational Behavior, 81*, 129-137.

Griffin, B., Loh, V., & Hesketh, B. (2013). A mental model of factors associated with subjective life expectancy. *Social Science & Medicine*, *82*, 79-86.

Grondin, S. (Ed.). (2008). *Psychology of time*. Emerald Group Publishing.

Groves, R. M., & Lyberg, L. (2010). Total survey error: Past, present, and future. *Public Opinion Quarterly*, *74*(5), 849 - 879.

Guilford, J. P. (1961). Factorial angles to psychology. *Psychological Review*, *68*(1), 1-20.

Guttman, L. (1952). Multiple group methods for common-factor analysis: Their basis, computation, and interpretation. *Psychometrika*, *17*(2), 209-222.

Haberman, S. J. & Dorans, N. J. (2011). Sources of score scale inconsistency. *Educational Testing Service, Research Report, 11-10.*

Halevy, Y. (2012). Time consistency: Stationarity and time invariance, *Working Paper.*

Hammond, K. R. (1948). Measuring attitudes by error-choice: an indirect method. *The Journal of Abnormal and Social Psychology, 43* (1), 38.

Hanisch, K. A. (1995). Behavioral families and multiple causes: Matching the complexity Of responses to the complexity of antecedents. *Current Directions in Psychological Science, 4* (5), 156 – 162.

Hanisch, K. A., & Hulin, C. L. (1990). Retirement as a voluntary organizational withdrawal behavior. *Journal of Vocational Behavior, 37*, 60-78.

Hayes, H. (2012). A generalized partial credit FACETS model for investigating order effects in self-report personality data. Unpublished doctoral dissertation. Georgia

Institute of Technology.

Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, *7*(3), 238-247.

Heidegger, M. (1927), Sein und Zeit. Halle: Max Niemeyer, (Translated as Being and Time by John Macquarrie and Edward Robinson. New York: Harper and Row, 1962.)

Hendrickson, A., Huff, K., & Luecht, R. (2010). Claims, evidence, and achievement-level descriptors as a foundation for item design and test specifications. *Applied Measurement in Education*, *23* (4), 358 -3 77.

Hershey, D. A. & Mowen, J. C., (2000). Psychological determinants of financial Preparedness for retirement. *The Gerontologist, 40(6),* 687-697.

Hershey, D. A. (2004). Psychological influences on the retirement investor. *Society of Certified Senior Advisors, 22*, 31 – 39.

Hershey, D. A., Henkens, K., & Dalen, H. P. (2010). Aging and financial planning for retirement: Interdisciplinary influences viewed through a cross-cultural lens.

Hershey, D. A., Jacobs-Lawons, J. M., McArdle, J. J., & Hamagami, F. (2007). Psychological foundations of financial planning for retirement. *Journal of Adult Development, 14,* 26-36.

Heisenberg, W. (1927) ‘Ueber den anschaulichen Inhalt der quantentheoretischen Kinematik and Mechanik’ *Zeitschrift für Physik, 43,* 172-198. English translation in (Wheeler and Zurek, 1983), pp. 62-84.

Hitt, M. A., Gimeno, J., & Hoskisson, R. E. (1998). Current and future research methods in strategic management. *Organizational Research Methods, 1*(1), 6-44.

Holtgraves, T. M. (2004). Social desirability and self-reports: Testing models of socially desirable responding. *Personality and Social Psychology Bulletin*, 30, 161 - 172.

Hom, P. Mitchell, T., Lee, T., & Griffeth, R. (2012). Reviewing employee turnover: Focusing on proximal withdrawal states and an expanded criterion. *Psychological Bulletin, 138*, 831-858.

Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology, 91* (3), 594 – 612.

Hurd, M. D., & McGarry, K. (2002). The predictive validity of subjective probabilities of survival. *The Economic Journal*, *112*(482), 966-985.

Ilmarinen, J., (2005). Towards a longer worklife! Ageing and the quality of worklife in The European Union. Helsinki: Finnish Institute of Occupational Health.

Jackson, D. N. (1971). The dynamics of structured personality tests. *Psychological Review, 78*(3), 229 - 248.

Jackson, D. N. & Messick, S. J. (1958). Content and style in personality assessment. *Psychological Bulletin, 55(4)*, 243 – 252.

Jacobs-Lawson, J. M., & Hershey, D. A. (2005). Influence of future time perspective, financial knowledge, and financial risk tolerance on retirement saving behaviors. *Financial Services Review-Greenwich-, 14*(4), 331.

James, W. (1890). *The consciousness of self. The principles of psychology,* 8.

Janssen, R., Tuerlinckx, F., Meulders, M., & De Boeck, P. (2000). A hierarchical IRT model for criterion-referenced measurement. *Journal of Educational and*

Behavioral Statistics, 25(3), 285 – 306.

Janssen, R., Schepers, J., & Peres, D. (2004). Models with item group predictors. In P. De Boeck & M. Wilson (Eds.), *Explanatory item response models: A generalized linear and nonlinear approach* (pp. 189-212). New York: Springer.

Janssen, R. (2010). Modeling the effect of item designs within the Rasch model. In Embretson, S. E. (Ed), (2010). *Measuring psychological constructs: Advances in model-based approaches*. (pp. 227-245). Washington, DC, US: American Psychological Association.

Jeon, M., Rijmen, F., & Rabe-Hesketh, S. (2013). Modeling differential item functioning using a generalization of the multiple-group bifactor model. *Journal of Educational and Behavioral Statistics*, *38*(1), 32-60.

Johnson, J. A. (1981). The “self-disclosure” and “self-presentation” views of item Response dynamics and personality scale validity. *Journal of Personality and Social Psychology, 40*, 761 – 769.

Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement, 38*, 319 – 342.

Kane, M. T. (2004). Certification testing as an illustration of argument-based validation. *Journal of Educational Measurement, 2* (3), 135 – 170.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50* (1), 1 – 73.

Kanfer, R., & Ackerman, P. L. (2004). Aging, adult development, and work motivation. *Academy of Management Review, 29*, 440-458.

Kaplan, D. (2000). Structural equation modeling: Foundations and extensions. Thousand

Oaks, CA: Sage

Kapteyn, A., Smith, J., Van Soest, A., & Vonkova, H. (2011). Anchoring vignettes and Response consistency. *Working Paper – RAND Labor & Population.*

Kastenbaum, R. (1961). The dimensions of future time perspective: An experimental analysis. *Journal of General Psychology, 65,* 203 – 218.

Kelly, G.A. *The Psychology of Personal Constructs.* New York: Norton, 1955.

Kerkhofs, M., & Lindeboom, M. (1995). Subjective health measures and state dependent reporting errors. *Health economics*, *4*(3), 221-235.

Kerry, M. J. & Hershey, D. A. (2015). Living to work and working to live: An empirical Review of the psychological antecedents and outcomes of retirement planning. Conditional acceptance for publication, in *The Journal of Work*, *Aging, & Retirement, 1*.

Kessler, E-M., & Staudinger, U. M. (2009). Afective experience in adulthood and old age: The role of affective arousal and perceived affective regulation. *Psychology and Aging, 24*(2), 349 –362.

Kim, S-H., Cohen, A. S., Alagoz, C. & Kim, S. (2007). DIF detection and effect size measures for polytomously scored items. *Journal of Educational Measurement, 44*(2), 93 – 116.

King, G., Murray, C. J., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review, 98*(1), 191-207.

Knowles, E. A. (1988). Item context effects on personality scales: Measuring changes the measure. *Journal of Personality and Social Psychology, 55,* 312-320.

Kohli, M., Rein, M., Guillemard, A-M., and van Gunsteren, H.(Eds) (1991). *Time for retirement: Comparative studies of early exit from the labour force*, Cambridge University Press, Cambridge.

Kooij, D. (2010). Motivating older workers – A lifespan perspective on the role of perceived HR practices. Unpublished dissertation. Vrije Universiteit.

Krause, N. (2001). Social support. In R. H. Binstock & George, L. K. (Eds.), *Handbook of Aging and the Social Sciences,* (272-294). San Diego, CA: Academic Press.

Kristensen, N., & Johansson, E. (2008). New evidence on cross-country differences in Job satisfaction using anchoring vignettes. *Labour Economics,15*(1), 96-117.

Krumboltz, J. D. (2009). The happenstance learning theory. Journal of Career Assessment, 17, 135-154.

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

Lance C. E., Baranik, L.E., Lau A. R., Scharlau, E. A. (2009). If it ain't trait it must be method: (mis)application of the multitrait-multimethod design in organizational research. In Statistical and Methodological Myths and Urban Legends: Doctrine, Verity, and Fable in the Organizational and Social Sciences, Eds. C. E. Lance & R. L. Vandenberg, pp. 337–60. New York: Routledge.

Lang, F. R. & Carstensen, L. L. (2002). Time counts: Future time perspective, goals, and Social relationships. *Psychology and Aging, 17*, 125 – 139.

Leary, L. F., & Dorans, N. J. (1985). Implications for altering the context in which test Items appear: A historical perspective on an immediate concern. *Review of Educational Research, 55* (3), 387 - 413.

LeBreton, J. W., Scherer, K. T., & James, L. W. (2014). Corrections for criterion reliability in validity generalization: A false prophet in a land of suspended judgment. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 7* (4).

Lee, H. W. (2011). An application of latent variable structural equation modeling for experimental research in educational technology. *The Turkish Online Journal of Educational Technology, 10*(1), 15-23.

Lee, R.M. (1993) Doing Research on Sensitive Topics. London: Sage.

Lee, T. W., Burch, T., & Mitchell, T. E. (2014). The story of why we stay: A review of Job embededness. *Annual Review of Organizational Psychology & Organizational Behavior, 1* (1), 199 – 216.

Lissitz, R. W. & Samuelsen, K. (2007) A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher, 36,* 437-448.

Luecht, R. M. (2006). Engineering the test: From principled item design to automated test assembly. Invited paper presented at the annual meeting of the *Society for Industrial and Organizational Psychology*, Dallas, Texas.

Lewin, K. (1951). *Field theory in social science; selected theoretical papers.* D. Cartwright (ed.). New York: Harper & Row.

Little, R. J., & Rubin, D. B. (1989). The analysis of social science data with missing values. *Sociological Methods & Research*, *18*(2-3), 292-326.

Löckenhoff, C. E. (2012). Understanding retirement: the promise of life-span Developmental frameworks. *European Journal of Ageing, 9,* 227 – 231.

Löckenhoff, C. E., O'Donoghue, T., & Dunning, D. (2011). Age differences in temporal discounting: the role of dispositional affect and anticipated emotions.*Psychology and Aging*, *26*(2), 274-284.

Locsin, R. C. (1993). Time experience of selected institutionalized adult clients. *Clinical Nursing Research*, 2(4), 451-463.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3* (S9) 635 – 694.

Lomranz, J., Shmotkin, D., & Katznelson, D. B. (1983). Coherence as a measure of future time perspective in children and its relationship to delay of gratification and social class. *International Journal of Psychology*, *18*(1-4), 407-413.

Lorge, I. (1937). The English Semantic Count. *Teachers College Record*, *39* (1), p. 65-77 http://www.tcrecord.org ID Number: 11956, Date Accessed: 6/1/2014 9:47:30 PM.

Luecht, R. M. (2006). Engineering the test: From principled item design to automated test assembly. Invited paper presented at the annual meeting of the *Society for Industrial and Organizational Psychology*, Dallas, Texas.

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading MA: Addison-Welsley Publishing Company.

Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin, 70*, 151 – 159.

MacCann, C., Ziegler, M., & Roberts, R. D. (2012). Faking in personality assessment. In Ziegler, M., MacCann, C. & Roberts, R. D. (Eds.), *New Perspectives on Faking in Personality Assessment,* 309 – 329. NY: Oxford University Press.

Maier, N. R. F. (1960). Maier's law. *American Psychologist*, *15*, 208–212.

Mantel, N., and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, *22*, 719-48.

Masche, J. G., & van Dulmen, M. H. M. (2004). Advances in disentangling age, cohort, and time effects: No quadrature of the circle, but a help. *Developmental Review*, *24*, 322 – 342.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods*, *44*(1), 1-23.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.

Mata, R. & Hertwig, R. (2011). Hot to model age-related motivational reorientations in Risky choice. Commentary on Depping and Freund. *Human Development, 54,* 368-375.

McVittie, C. & Goodall, K. (2012). The ever-changing meanings of retirement. *American Psychologist, 76*(1), 75-77.

Meade, A. W. (2010). A taxonomy of effect size measures for the differential functioning of items and scales. *Journal of Applied Psychology, 95*(4), 728 – 743.

Meehl,P. E.(1945). Reprinted with Prefatory Comment in L. D. Goodstein & R. I. Lanyon (Eds.), Readings in personality assessment (pp. 245–253). New York: Wiley, 1971

Meehl, P. E. (1967). Theory-testing in psychology and physics: a methodological paradox. *Philosophy of Science, 34,* 103 – 115.

Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry, 1*(2), p. 108 – 141.

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research, 13*, 127-143.

Messick, S. & Jackson, D. N. (1958). Content and style in personality assessment. *Psychological Bulletin, 55(4)*, p.243 - 252.

Messick, S. (1989).V alidity. In R. L. Linn(Ed.),E ducational measurement (3rd ed., pp.13-103). New York: American Council on Education/ Macmillan.

Messick, S. (1995). Validity of psychological assessment. Validation of inferences from person's responses and performances as scientific inquiry into score meaning. *American Psychologist, 50* (9), 741 – 749.

Millar, M.G., & Tesser, A. (1986). Effects of affective and cognitive focus on the attitude-behavior relation. *Journal of Personality and Social Psychology, 51*, 270 – 276.

Millshap, R. E. & Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement, 17* (4), 297 – 334.

Meyer, R. D., Dalal, R. S., & Hermida, R. (2010). A review and synthesis of situational Strength in the organizational sciences. *Journal of Management, 36(1)*, 121-140.

Montalto, C. P., Y. Yuh & S. Hanna (2000). Determinants of planned retirement age, *Financial Services Review, 9,* 1-15.

Mowen J. C. (2000). The 3M model of motivation and personality: theory and empirical applications to consumer behavior. Kluwer Academic Publishers.

Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement*, *14*(1), 59-71.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied psychological measurement*, *16*(2), 159-176.

Nolet, V. & Tindal, G. (1990). Evidence of construct validity in published achievement tests. Paper presented at the *Annual Meeting of the American Educational Research Association*: Boston, MA.

Ng, T. W. H., Feldman, D. C. (2007). Organizational embededness and occupational embededness across career stages. Journal of Vocational Behavior, 70, 336-351.

Ng, T. W. H., & Feldman, D. C. (2008). The relationship of age to ten dimensions of job performance. Journal of Applied Psychology, 93, 392-423.

Noone, J. H., Stephens, C., & Alpass, F. (2010). The process of retirement planning scale (PRePS): Development and validation. Psychological Assessment, 22(3), 520 – 531.

Nuttin, J. R. (1964). The future time perspective in human motivation and learning. *Acta Psychologica*, *23*, 60 – 82.

Nuttin, J. R. (1973). Time perspective: theoretical and methodological considerations. *Psychological Belgica, 13*(3), 265-294.

Nye, C. D. & Drasgow, F. (2011). Effect size indices for analyses of measurement equivalence: Understanding the practical importance of differences between groups. *Journal of Applied Psychology, 96*(5), 966 – 980.

Nye, C. D., Newman, D., & Joseph, D. L. (2009). Never say “always”? Extreme item wording effects on scalar invariance and item response curves. *Organization*

Research Methods, 13 (4), 806 – 830.

Öncel, L. (2014). Career Adapt-Abilities Scale: Convergent validity of subscale scores. *Journal of Vocational Behavior*, *85*(1), 13-17.

Paulhus, D. L. (2002). Socially desirably responding: The evolution of a construct. In H.

I. Brown, D.N Jackson, & D.E. Wiley (Eds.), The Role of Construct in Psychological and Educational Measurement (pp. 49 – 69). Mahwah: NJ: Erlbaum.

Paglieri, F. (2013). The costs of delay: Waiting versus postponing in intertemporal choice. *Journal of the Experimental Analysis of Behavior, 99,* 362-377.

Payne, J. W., Sagar, N., Shu, S. B., Appelt, K. C., & Johnson, E. J., (2013). Life expectancy as a constructed belief: Evidence of a live-to or die-by frame. *Journal of Risk and Uncertainty, 46,* 27-50.

Pearl, J. (2011). Invited commentary: Understanding bias amplification. *American Journal of Epidemiology, 174* (11), 1223 – 1227.

Pearl, J. & Bareinboim, E. (2012). External validity: From do-calculus to transportability across populations. *Statistical Science, R-400*.

Penfield, R. D. (2014). An NCME instructional module on polytomous item response Theory models. *Educational Measurement: Issues & Practice*, *33*(1), 36 – 48.

Platt, J. R. (1964), Strong inference. *Science, 146,* 347-53.

Podsakoff, P. M. & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management, 12*(4), 531 – 544.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual*

Review of Psychology, 65, 539 - 569.

Pöppel E. (1989) Taxonomy of the subjective: an evolutionary perspective. In: *Neuropsychology of Visual Perception* (Ed. J. W. Brown). Lawrence Erlbaum Associates, Hillsdale, p. 219-232.

Reckase, M. D. & McKinley, R. L. (1991). The dis-criminating power of items that Measure more than one dimension. *Applied Psychological Measurement*, *15*, 361 - 373.

Roeckelein, J. E. (2000). *The concept of time in psychology: A resource book and annotated bibliography*. Greenwood Publishing Group.

Rozema, L. A., Darabi, A., Mahler, D. H., Hayat, A., Soudager, Y., & Steinberg, A. M. Violation of Heisenberg's measurement-disturbance relationship by weak measurements. *Physical Review Letters, 109.*

Sackett, D. L. (1979). Bias in analytic research. *Journal of chronic diseases*, *32* (1), 51-63.

Sadana, R., Mathers, C. D., Lopez, A. D., Murray, C. J. L., & Iburg, K. (2000). Comparative analysis of more than 50 household surveys on health status, GPE Discussion Paper No 15, World Health Organisation, Geneva.

Sargent, L., Lee, M., Martin, B., & Zikic, J. (2013). Reinventing retirement: New pathways, new arrangements, new meanings. *Human Relations*, *66* (1), 3 – 21.

Schaie, W. K. (1965). A general model for the study of developmental problems. *Psychological Bulletin, 64* (2), 92 – 107.

Schoen, J. L., DeSimone, J. A., James, L. R. (2011). Exploring joint variance between independent variables and a criterion: Meaning, effect, and size. *Organizational*

Research Methods, *14*, 674 - 695.

Schwarz, N. (1999). Self-reports: How the questions shapes the answers. *American Psychologist, 54* (2), 93 – 105.

Shafer, G. (1976). *A mathematical theory of evidence* (Vol. 1). Princeton: Princeton university press.

Shepard, L. A. (1993). Evaluating test validity. In L. Darling-Hammon (Ed.), *Review of Research in Education*, Vol. 19. Washington DC: AERA.

Shultz, K. S. & Olson, D. A. (2013). The changing nature of work and retirement. In *The Oxford Handbook of Retirement*, pp. 543 – 558.

Shultz, K. S. & Wang, M. (2011. Psychological perspectives on the changing nature of retirement. *American Psychologist*, *66* (3), 170 – 179.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63* (2), 129 - 138.

Sloper, D. (2006 – 2014). Statistics calculators. Retrieved on May 5, 2014 from http://www.danielsoper.com/statcalc3/.

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology, 27*, 229 – 239.

Spector, P. E., Brannick, M. T. (2011). Methodological urban legends: The misuse of statistical control variables. *Organizational Research Methods, 14*, 287 - 305.

Steiger, J.H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*, 173 – 180.

Stone, A. A., Turkkan, J. S., Bachrach, C. A., Jobe, J. B., Kurtzman, H. S., & Cain, V. S.

(2000). *The Science of Self-Report: Implications for Research and Practice*.

Suen, H. K. (1990). *Principles of test theories.* Hillsdale, N.J. Earlbaum.

Suppes, P. (1968). The desirability for formalization in science. *Journal of Philosophy, 65,* p.651– 664.

Thorndike, R. L. (1947). Research problems and techniques. *Army Air Forces, 3*: Washington DC Psychology Program.

Thurstone, L. L. *Vectors of Mind: Multiple-Factor Analysis for the Isolation of Primary Traits*. (1935). University of Chicago Press, Chicago: Illinois.

Tourangeau, R. & Rasinski, K. A. (1988). Cognitive processes underlying context effects In attitude measurement. *Psychological Bulletin, 103* (3), 299 – 314.

Treadway, D. C. Duke, A. B., Perrewe, P. L., Breland, J. W., & Goodman, J. M. (2012). Time may change me: The impact of future time perspective on the relationship between work-family demands and employee commitment. *Journal of Applied Social Psychology, 41*(7), 1659-1679.

Trommsdorff, G. (1983). Future orientation and socialization. *International Journal of Psychology, 18*, 381 - 406.

Tversky, A., & Koehler, D. J. (1994). Support theory: a nonextensional representation of subjective probability. *Psychological Review*, *101*(4), 547-567.

van Alphen A., Halfens R., Hasman A., Imbos T. (1994). Likert of Rasch? Nothing is more applicable than good theory. *Journal of Advanced Nursing*, 20, 196 – 201.

Van der Ark, A. L. (2001). Relationships and properties of polytomous item response Theory models. *Applied Psychological Measurement, 25*, 273-282.

van der Maas, H. L. J., Molenaar, D., Maris G., Kievit R. A. & Borsboom, D. (2011).

Cognitive psychology meets psychometric theory: On the relation between process models for decision making and latent variable models for individual differences. *Psychological Review, 118,* 339 - 356.

van Solinge, H. & Henkens, K., (2010). Living longer, working longer? The impact of subjective life expectancy on retirement intentions and behavior. *European Journal of Public Health, 20(1)*, 47-51.

Viswanathan, M. (2005). *Measurement Error and Research Design.* Sage Publications.

Wakker, P. P. (2004). On the composition of risk preference and belief. *Psychological review*, *111*(1), 236-255.

Wallace, M. (1956). Future time perspective in schizophrenia. *Journal of Abnormal Psychology,52*, 240 – 245.

Wallace, M. and Rabin, A.I. (1960). Temporal experience, *Psychological Bulletin, 57*, 213 - 236.

Wand, J., King, G., & Lau, O. (2011). Anchors: Software for anchoring vignettes data. *Journal of Statistical Software 42*(3), 1-25.

Wang, M. & Schultz, K. S. (2010). Employee retirement: A review and recommendations For future investigation. *Journal of Management, 36* (1), 172 – 206.

Wang, M. & Shi, J. (2014). Psychological research on retirement. *Annual Review of Psychology, 65* (1), 1-25.

Wang, W. C., Shih, C. L., & Sun, G. W. (2012). The DIF-free-then-DIF strategy for the assessment of differential item functioning. *Educational and Psychological Measurement*, *72*(4), 687-708.

Windschitl, P. D. (2000). The binary additivity of subjective probability does not indicate the binary complementarity of perceived certainty. *Organizational Behavior and Human Decision Processes*, *81*(2), 195-225.

Winkel, D. E., Wyland, R. L, Shaffer, M. A., & Clason, P. (2011). A new perspective on psychological resources: Unanticipated consequences of impulsivity and emotional intelligence. *Journal of Occupational and Organizational Psychology*, *84*, 78-94.

Yang, T. Y. & DeVaney, S. A., (2011). Intrinsic rewards of work, future time perspective, the economy in the future, and retirement planning. *The Journal of Consumer Affairs*, *45*(3).

Zacher, H., & Frese, M. (2009). Remaining time and opportunities at work: Relationships between age, work characteristics, and occupational future time perspective. *Psychology and Aging*, *24*(2), 487-493.

Zacher, H., Heusner, S., Schmitz, M., Zwierzanska, M. M., & Frese, M. (2010). Focus on opportunities as a mediator of the relationships between age, job complexity, and work performance. Journal of Vocational Behavior, 76(3), 374-386.