REPORT

Measuring Ageism: A Literature Review

DATE
JANUARY 24, 2019

PRESENTED TO:
AARP

PRESENTED BY:
Academic Research Centers
NORC at the University of Chicago
Michele Zimowski, PhD
Kevin L. Brown, PhD
Christopher Cook
This report was created with support
from the AARP.

NORC
at the UNIVERSITY *of* CHICAGO

Table of Contents

Age-ism describes the subjective experience implied in the popular notion of the generation gap. Prejudice of the middle-aged against the old in this instance, and against the young in others, is a serious national problem. Age-ism reflects a deep seated uneasiness on the part of the young and middle-aged—a personal revulsion to and distaste for growing old, disease, disability; and fear of powerlessness, "uselessness," and death.

–Butler (1969, p. 243)

Introduction and Overview

The term "ageism" was coined in 1969 during a tumultuous period in American history when issues of class and race had come to the forefront of national politics. At the time examples of age discrimination in public housing, for instance, may have seemed less critical than discrimination on the basis of racial or class membership. However, Butler already understood that ageism intersects with racism and classism to compound the problem, and the massive demographic shift since 1969 toward an older population means that discrimination against the elderly has rapidly become a social problem of equal standing, requiring focused attention from academics and policymakers alike (Palmore 2001). Indeed, studies of public attitudes towards older adults had accumulated sufficiently just 16 years after Butler coined the term to merit a meta-analytic review (Kite and Johnson 1988). Research on the topic only grew from there, with an updated review by Kite and her coauthors published in 2005 and, more recently, publication of a special issue of the *Journal of Social Issues* (2016) focused on ageism in health and employment contexts. It should be noted that the great majority of this research addresses what Butler (1980) called "malignant" ageism (stereotyping or bias) rather than "benign" ageism (anxiety or fear of aging).

Ageism and the Mass Media

A number of theories have been offered to explain the rise of negative attitudes towards the elderly, particularly in America where a growing "culture of youth" has come to predominate basic social institutions, particularly the media (Berger, 2017). "Common feelings toward the elderly are that they are of low status and incompetent," writes Berger (2017, p. 184). Whatever its origins, he says that research has consistently shown that the media reinforce cultural stereotypes by portraying the elderly as "frail, feeble, financially distressed, and not contributing to society" (ibid). These stereotypical messages, in turn, affect both how the elderly are viewed and how older people feel about themselves (Robinson, Gustafson, and Popovich, 2008). Importantly, this internalization of negative age stereotypes impacts the health and functioning of older Americans (Levy, 2009), while opportunities for internalizing such stereotypes are legion. A 1993 review of the literature on the portrayal of the elderly in the media found evidence of negative age stereotyping across media sources, including the full range of television programming with the possible exceptions of daytime soap operas and cartoons. Conversely, children's literature and picture books did show consistent age stereotypes across studies, while results from various

print media for adults such as magazines and newspapers yielded mixed results.[1] Hollywood films also tend to reinforce stereotypes by providing inaccurate pictures of older adults and numerous examples of ageist comments (Smith, Pieper, and Choueiti, 2018).

A more recent review of research agreed that portrayals of older people in mass media are generally negative but expanded on this conclusion by noting that such bias can be more subtle than direct stereotyping (Harwood, 2007). For example, older adults may be well-represented in advertising and even portrayed in positive or "desirable" ways (Miller, Leyell, and Mazachek, 2004), but the great majority of these advertisements are for health products, reinforcing the stereotypical association between aging and poor health. Similarly, older adults may appear in reasonable numbers but usually have peripheral roles in television shows and movies and often serve as the butt of jokes in comedy programs or even in birthday cards. Harwood (2007, p. 174) notes that there are plenty of examples of positive images of aging in the media, but on balance the "messaging" is sufficiently negative to sustain cultural biases and instill ageist attitudes. Indeed, negative images of the elderly in the long-running TV show *The Simpsons* (Blakeborough, 2008) to the perpetuation of stereotypes via social media platforms such as Facebook (Levy et at., 2013) show that "ageism is a significant social problem," writes Berger (2017). "Researchers and academics need to place more emphasis on this larger cultural problem and try to figure out ways to minimize the impact on our aging population" (p. 185).

Overview of Ageism Research

Although much research on ageism "has shown that by the time people reach their college years they have already developed a deeply ingrained perspective on the elderly and have internalized dominant cultural assumption about the elderly" (Berger 2017, p. 185), some studies have found a mix of positive and negative attitudes. Kite and Johnson's 1988 meta-analysis of the ageism literature concluded that 30 of 43 studies found that people have more negative attitudes towards older than younger people, though the differences (effect sizes) reported in each study varied widely. They attributed this variation in findings to the influence of "moderators" such as social context (e.g. providing specific versus general information about an older person) or the dimensions of attitudes being assessed (e.g. competence versus personality traits). The updated meta-analysis by Kite and colleagues (2005) included many more studies and thus was able to explore the influence of these moderators in more detail. Overall, their review included 232 individual studies and found significant bias towards older versus younger people in the aggregate, with an average effect size of 0.24 for the 100 studies that focused on "psychological tendencies that are expressed by evaluating a particular entity with some degree of favor or disfavor" (Kite et al., 2005, p. 244). Studies that considered other dimensions of ageism such as holding ageist

---

[1] Note that findings are more consistent when considering the *representation* of older people in electronic and print media, with 20 out of 22 studies agreeing that the elderly are underrepresented relative to their proportion of the U.S. population. Lauzen and Dozier (2005) and a series of more recent studies funded by the USC Annenberg Inclusion Initiative added top-grossing films to the list of media sources that underrepresent older adults, particularly women. Robinson and Skill (1995) found that portrayals of the elderly on television had not responded to the changing demographic situation, although more recent research shows that older adults are now beginning to see the long-expected decline in such "visual ageism" (Loos and Ivan 2018).

stereotypes (n=11) and assessing competence (n=75) or physical attractiveness (n=17) yielded even larger effect sizes. More fine-grained analysis showed that effect sizes were moderated by:

■ *Information about target*. More bias was evident when the respondent was provided little to no detailed information about the person being evaluated, which the authors suggest supports a "social role theory" approach to studying ageism.

■ *Gender of target*. As expected, age bias tends to be greater when women are being assessed compared to men except when assessing the competence of targets, where the age gap was larger for males than females.

■ *Characteristics of respondent*. Middle-aged respondents showed a greater preference for younger adults in studies that assessed bias or competence. Little difference was seen in effect sizes across respondent populations such as undergraduates, professionals, or the general population.

■ *Study design*. Studies that assessed the same respondents under different treatments or conditions (within-subject design) found considerably larger perceived bias against older adults across domains than did studies that used between-subject designs.

Overall, the 2005 meta-analysis confirmed and reinforced the findings of the 1988 study while providing additional insight into how findings can vary (or be moderated) by the characteristics of the respondent sample, target population, and study design. A more recent assessment of the literature (Levy and Macdonald, 2016) suggests additional avenues of research, including a more intentional classification of the psychological dimensions of ageism and a reassessment of the tools being used to measure them. "To aid in the expansion of research in this area," Levey and MacDonald (2016) suggest "studying ageism as a tripartite attitude…with a stereotyping or cognitive component (e.g., associating positive and negative attributes), a prejudice or affective component (e.g., liking or disliking, valuing or devaluing), and a discrimination or behavioral component" (p. 13). The remainder of this review uses this tripartite approach to organize our assessment of the measures that have been employed to study ageism over the last 50 years.[2]

Review Methods and Approach

Relevant literature was identified and accessed via Web searches in November-December 2018. Searches used several online tools, including Google Scholar, PubMed, PsycINFO, and federal grants databases. As shown below in Exhibit 1, online searching was broken into two phases – an initial phase that sought to identify major instruments used by researchers to study ageism, and a subsequent phase that focused on applications and assessments of the instruments identified. In the first phase, search terms incorporated two elements: (1) a term evoking the phenomenon of ageism, and (2) a term evoking scientific methods for detecting or quantifying the phenomenon. In the second phase, search terms consisted primarily of the names of the relevant instruments surfaced in Phase 1.

---

[2] Note that we are focusing here on psychological measures and not measures of discriminatory behaviors such as preferences in hiring or observed interactions. We also say little here about using measures such as eye-tracking (e.g. Fung et al. 2015) or subliminal messaging (e.g. Levy 2009) given their limited applications to-date.

Exhibit 1: Key Search Terms Used to Identify Measures of Ageism

Phase 1 search terms and phrases	Phase 2 search terms and phrases
■ Ageism measure ■ Ageism scale ■ Ageism instrument ■ Agism measure ■ Agism scale ■ Agism instrument ■ Ageist measure ■ Ageist scale ■ Ageist instrument	■ Kogan's Scale ■ Aging Semantic Differential ■ Facts on Aging Quiz ■ Expectations Regarding Aging Scale ■ Anxiety about Aging Scale ■ Image of Aging Scale ■ Ambivalent Ageism Scale ■ Fraboni Ageism Scale ■ Succession, Identity, and Consumption Ageism Scale ■ Braithwaite's Scale ■ Relating to Older People Evaluation

Instruments were selected for inclusion based on their relevance to measuring ageism or a closely related phenomenon. We focused on instruments with contemporary salience; this includes some measures developed in the last two years, but also measures developed over half a century ago that continue to be widely used and discussed in the field. We excluded measures such as the Attitudes to Ageing Questionnaire that did not seem appropriate for conducting research on interventions to combat ageism in younger people and also focused primarily on peer-reviewed publications given the highly technical issues involved in assessing the psychometric properties of measurement tools.

I. Measuring Ageist Stereotypes

The largest group of measures used to study ageism include the earliest and focus primarily on the first of the three elements of the theorized "tripartite" ageist attitude, namely the stereotyping or cognitive component.

Kogan Attitudes Toward Old People Scale

Developed almost a decade before Butler (1969), the Kogan Attitudes Toward Old People Scale was developed as a scientific tool for measuring attitudes toward older people. In this era, social scientists were investigating older people as a sort of "quasi-minority," who faced experiences of stereotyping and discrimination analogous to those of ethnic and religious minorities (Barron, 1953), but who diverged from these other minority groups in that they did not function as an independently-functioning cultural sub-group (Drake, 1959). Kogan (1961) cited the work of Tuckman and Lorge (1953) in providing empirical evidence of stereotyping and misconceptions about older people, but pointed out that they "make no use of attitude scaling procedures and pay little attention to psychological correlates of attitudes toward old people." Kogan's OP Scale consists of 17 pairs of statements – one positive statement and one negative statement in each pair – about older people [e.g. "It would probably be better if most old people lived in residential units with people of their own age." "It would probably be better if most old people lived in residential units that also housed younger people." (Kogan, 1961)]. In addition to residence, the

topics addressed include: adaptability, relations, dependence, personality, cognitive capacity, political power, and appearance. Respondents rate their agreement with these statements on a Likert scale, thereby allowing investigation of attitudes toward older people with relation to social norms and individual differences.

Kogan's Scale has been used and discussed widely by gerontological scholars over the last half-century. In the late 1990s, Hilt and Lipschultz (1999) raised reasonable concerns about the usage of Kogan's Scale in contemporary society, including the length of the scale and its use of outdated expressions. To address these issues, Hilt developed a shorter version; however, due to concerns about the rationale for adjusting the instrument, the small sample sizes in Hilt's study, and problematic reliability, scholars continue to prefer and employ Kogan's original scale (Iwasaki & Jones, 2008). The OP tool consists of 17 matched pairs of positive and negative statements. The set of positive items and the set of negative items were intended to serve as two parallel forms of the instrument. The OP asks respondents to indicate the extent to which they agree or disagree with each statement on a six-point scale—strongly disagree, disagree, slightly disagree, slightly agree, agree, and strongly agree. Exhibit 2 displays six statements representing three of the 17 matched pairs in the tool.

Exhibit 2: Select Items in the Kogan Attitudes toward Old People (OP) Scale

Type of Statement	
Negative	**Positive**
It would probably be better if most old people lived in residential units with people of their own age.	It would probably be better if most old people lived in residential units that also housed younger people.
Most old people would prefer to quit work as soon as pensions or their children can support them	Most old people would prefer to continue working just as long as they possibly can rather than be dependent on anybody.
It is foolish to claim that wisdom comes with old age.	People grow wiser with the coming of old age.

The tool was administered to two samples (N=128 and N=186) of male undergraduate students at Northeastern University and one sample of 168 male and female undergraduate students at Boston University to examine the measurement properties of the tool. Responses to the items were summed across the negative items in the tool (O- items) and across the positive items in the tool (O+ items) to generate scores on the O- and O+ subscales. The reliability of scores ranged from 0.73 to 0.83 on the negative scale (O-) and from 0.66 to 0.73 on the positive subscale (O+) across the three samples. The correlations between the scores on the two subscales (purportedly representing parallel forms of the OP) ranged from 0.46 to 0.51 across the three samples, indicating that the scores on the subscales shared 26 percent of their variance at best.[3]

In 2008, Iwasaki and Jones conducted a study in which they examined the psychometric properties of the OP in a sample of 512 students enrolled in undergraduate courses at a state university in the Midwest.

---

[3] Note that responses to the positive items are reverse-coded so that higher scores on both subscales reflect more unfavorable attitudes toward old people.

The reliabilities obtained for the negative and positive subscales, 0.79 and 0.68, respectively, were in line with the earlier results of Kogan (1961). Like Kogan (1961), Iwasaki and Jones (2008) found that scores on the two subscales shared only 26 percent of their variance. Iwasaki and Jones (2008) also looked at item-total score correlations and found that a number of the items exhibited low correlations with the total score, indicating that several items in the instrument are, at best, only weakly related to the construct they are intended to assess. In addition, they found non-significant or relatively small correlations between the positive and negative items in 11 of the 17 matched item pairs. That many of the items appear outdated or poorly worded could account for these less-than-favorable findings.

The OP is one of the first instruments developed to measure positive and negative attitudes toward old people, and one of the most widely used measures of ageism. *Even so, examinations of the measurement properties of the OP have relied almost exclusively on samples of undergraduate students. Consequently, little is known about how the OP operates among individuals of other ages. Although the positive and negative items in the OP were designed to serve as parallel forms of the instrument, the correlations between the individual items in the matched pairs and those between the O- and O+ subscales indicate that the tool is not operating as intended.* It is not clear whether the lack of a strong relationship between the two subscales is simply due to poorly performing items in one or both subscales, or whether it indicates that the positive and negative items are measuring somewhat different constructs. The tool would benefit from further work to disentangle these sources of variance in the scores. Of note, the reliabilities of the subscales are lower than those typically found for scales of similar length. That a number of items exhibit low item-total score correlations may be contributing to the less than striking reliabilities found for the O- and O+ subscales.

Aging Semantic Differential (ASD)

The Aging Semantic Differential (ASD) is perhaps the most widely used instrument in educational contexts for assessing stereotyping attitudes toward older adults (Gonzales, Tan, & Morrow-Howell, 2010). Rosencranz and McNevin (1969) developed the instrument based on the semantic differential research of Osgood, Suci, and Tannenbaum (1957), who suggest that a full understanding of variance in judgments must take into account three judgment factors: evaluative, potency, and activity. Rosencranz and McNevin (1969) adapted this model to study attitudes toward older adults by proposing that these factors could be equated to spectrums of *personal acceptability-unacceptability* (evaluative), *instrumental-ineffective* (potency), and *autonomous-dependent* (activity). The instrument consisted of 32 contrasting descriptors (e.g. pleasant-unpleasant; strong-weak; decisive-indecisive), each accompanied by a 7-point scale indicating likeness to one adjective or its opposite. Respondents were asked to mark the scale at a point which best described their judgment of men aged 20-30, 40-55, and 70 or older.

Following decades of debate which called into question the ASD's three-factor structure, validity, and use of outdated language (Polizzi & Steitz, 1998), Polizzi developed a refined ASD. This version used more contemporary adjectives and shifted from a three-factor model to a single-factor model, focusing only on the evaluative sphere that Rosencranz and McNevin had called *personal acceptability-unacceptability*. The latter change was predicated on evidence from Osgood's and Suci's (1955) and Polizzi's (2003) work suggesting that the evaluative factor was responsible for a large proportion, perhaps even a majority, of

the variance captured. In its original form, the ASD consisted of 32 polar adjective pairs, each with a seven-point semantic differential scale. The positive adjective of each pair appeared on the right of the accompanying scale, the negative adjective on the left of the scale. Respondents were asked to place a check mark at the point along each scale that best described their view of men age 20 to 30, age 40 to 55, and age 70 or older. Exhibit 3 displays adjective pairs representative of those in the original ASD.

Exhibit 3: Select polar adjective pairs from the original ASD

Polar Adjective Pairs	
Positive	**Negative**
Progressive	Old-fashioned
Consistent	Inconsistent
Healthy	Unhealthy
Tolerant	Intolerant
Pleasant	Unpleasant
Flexible	Inflexible
Neat	Untidy
Trustful	Suspicious

Considerable attention has been paid to the measurement properties of ASD. Much of the effort has been devoted to determining the factor structure of the ASD, with little thought given to other aspects of the instrument. An exception to that rule is the work of Polizzi (2003). In 2003, he pointed out that one of the major problems with the ASD lies in its outdated adjectives that may no longer apply to attitudes toward the elderly as those attitudes exist today. He also noted that the development of the instrument relied exclusively on men as the attitudinal objects, even though attitudes toward elderly men and women might not be the same. To address these concerns, Polizzi (2003) set out to develop a version of instrument with updated adjective pairs and then to test it with elderly men as well as elderly women serving as the attitudinal objects. For purposes of developing a revised version of the ASD, Polizzi assembled a list of 81 adjective pairs, 30 of which were drawn from the original ASD and then updated (e.g., handsome-ugly changed to attractive-unattractive, neat-untidy changed to neat-messy), the others from a range of sources, including Osgood and Suci's original list of adjective pairs. Three-hundred undergraduate students were asked to respond to a version of the ASD that included these 81 adjective pairs. The students were asked to complete the instrument twice, once when the attitudinal object was a "man 70-85 years of age" and once when the attitudinal object was a "woman age 70-85." Polizzi (2003) performed a number of factor analyses on the data to examine the factor structure of the 81 items in both cases. He found strong similarities between the factor structures when the attitudinal object was an elderly woman and when it was an elderly man, noting that the ordering of the loadings on the factors differed somewhat across the two solutions.

Based on the results of the analyses, Polizzi (2003) arrived at a revised version of the ASD consisting of 24 adjective pairs. A few years later, Iwasaki and Jones (2008) administered Polizzi's revised version of

the ASD to a sample of 785 undergraduates with "older adults" serving as the attitudinal object. The revised version of the ASD performed well in the sample yielding an alpha coefficient of 0.93 and item-total score correlations ranging in value from 0.45 to 0.64. To evaluate the factor structure of the instrument when older adults served as the attitudinal object, Iwasaki and Jones (2008) split their sample into two random halves, performing an exploratory factor analysis on one half of the data and a confirmatory factor analysis on the other half. After noting that the first four eigenvalues of the correlation matrix were 9.38, 1.71, 1.31, and 1.03, respectively, Iwasaki and Jones went on to fit one-, two-, three-, and four-factor models to the data in the exploratory analysis, examining orthogonal as well as oblique solutions. Based on the results of the exploratory analysis, a number of models were fit to the data in the confirmatory stage. Several indices were generated to evaluate the fit of the models.

Although Iwasaki and Jones (2008) concluded that the one-factor model was clearly an improvement over the null model, they also suggested that the models including correlated factors appeared to fit the data better even though none of the multi-factor confirmatory models fit the data particularly well. *By all indications, however, Pollizzi's (2003) version of the ASD is a strong measure of the attitudes young people have toward older adults. It consistently exhibits alpha coefficients above 0.90 and a dominant first factor that accounts for nearly 40 percent of the variance in the scores regardless of the attitudinal object under study.*

The Facts on Aging Quiz (FAQ)

In 1976, Erdman Palmore developed the Facts on Aging Quiz to demonstrate misconceptions about aging and older people among students at Duke University, where he was teaching a course on the social aspects of aging (Palmore, 1998). The 25-item quiz consists of a series of factual statements about aging and requires the respondent to judge whether each statement is true or false. After discovering that no similar short, objective survey existed in the literature, and noting confusion surrounding widely-used instruments such as Kogan's OP Survey (Kogan, 1961), Palmore refined and published the FAQ1 (Palmore, 1977). Over the following decades, Palmore's quiz came to be widely used in both education and research and was traditionally administered to groups such as students, teachers, and health care professionals (Palmore, 1998). Additionally, Palmore and others continued to test and refine the quiz over several years. The original FAQ tool consisted of 25 true and false statements designed to measure basic factual knowledge about old age and aging. Exhibit 4 displays a few of the statements in the quiz.

Exhibit 4: Select statements from the Facts on Aging Quiz (FAQ)

All five senses tend to decline with old age.
Older people usually take longer than young people to learn something new.
Older workers tend to have fewer accidents than younger workers.
The majority of medical practitioners give low priority to the aged.

A few years after developing the FAQ, Palmer (1981) introduced an alternative version of the tool (FAQ2) that, like the first FAQ, requires a true or false response to each of 25 statements. Both versions of the tool have received considerable attention in the literature and have undergone any number of revisions in an effort to improve upon the measurement properties of the tool. While some researchers have revised the wording of the items, others have added a "don't know" response option to the tool to discourage guessing. Still others have converted the statements into multiple-choice questions to reduce the impact of guessing on the scores. In spite of these efforts, both versions of the tool consistently yield reliability estimates in the unacceptable range (below 0.60). *Thus the FAQ in its current form is not suitable for measuring basic factual knowledge about old age and aging due to its poor reliability. Attempts to improve upon measurement properties of the tool have been unsuccessful thus far.*

The Expectations Regarding Aging (ERA) Scale

The Expectations Regarding Aging Scale emerged from a line of research investigating the connection between negative expectations about the aging process and negative health outcomes in older adulthood. In 1981, Kart (1981) suggested that "Over-attribution of symptoms to the aging process directs the attention of the elderly person away from real disease and/or environmental factors that may affect health. Such misattributions may have tragic consequences." In the following decades, researchers produced empirical evidence that attribution of health conditions to aging is associated with greater acceptance of symptoms, decreased use of preventive health measures, and increased mortality. Sarkisian et al (2002) responded by developing the Expectations Regarding Aging Scale (ERA-38) in order to equip researchers with a valid instrument to measure older adults' perceptions about the aging process, with the goal of examining the relationship between expectations regarding aging, health behaviors, health outcomes, and service use. The tool consists of 38 statements covering 10 domains of aging. Statements in the first section of the tool address expectations regarding the respondent's own aging. Statements in the second section of the tool address expectations regarding the aging of older adults in general. Within each type of expectation, the tool asks respondents to select one of four response options (definitely true, somewhat true, somewhat false, and definitely false) that best describes their feelings about the statement. Exhibit 5 displays one statement representing one or the other type of expectation for each of the 10 domains covered in the instrument.

Exhibit 5: Select Items from the Expectations Regarding Aging (ERA-38) scale

Type of Expectation	Domain	Sample Statement
Expectation regarding their own aging	Cognitive Function	I expect that as I get older I will become more forgetful.
	Mental Health	I expect that as I get older I will enjoy my life.
	Sexual Function	I expect that as I get older my sexual desire will decrease.
	Fatigue	I expect that as I get older I will get tired more quickly.
	Appearance	I expect that as I get older I will become less attractive.
	Functional Independence	When I get older I expect I will be able to do everything I want to.

Expectation regarding the aging of older adults in general	General Health	When people grow older, one thing or another is going to go wrong with their body.
	Pain	Having more aches and pains is an expected part of getting old.
	Urinary Incontinence	Needing to use adult diapers is just an expected part of getting old.
	Sleep	It's a normal part of aging that older people have trouble sleeping.

Development of the ERA-38 involved several stages. To identify relevant domains to cover in the instrument, Sarkisian et al. (2002) held focus groups and conducted one-on-one interviews with over 40 older adults. Based on the domains identified, draft versions of 94 items were developed, revised through cognitive interviews, and then field-tested in a sample of 58 older adults. Poorly performing items were eliminated from the pool and the 56 remaining items were administered to a sample of 429 older adults, along with a battery of other instruments, to develop and validate the final version of the tool. After items displaying low item-test correlations, ceiling effects, or redundancies with other items were eliminated from the pool, 38 items remained in the tool. All subsequent analyses evaluating the validity and reliability of the ERA-38 were based on those 38 items and the data collected from the 429 older adults who participated in this final stage of the tool's development. Exhibit 6 shows the distribution of the ERA-38 items across the 10 domains, along with estimates of the internal consistency of the domain/subscale scores and the total score on the scale. The number of items representing each domain varied from 1 to 12 items. The alpha coefficients ranged in value from 0.58 (Pain) to 0.89 (Mental Health).[4] Of note, the reliabilities found for the subscales with five or fewer items are considerably higher than those typically found for scales of similar length. With a value of 0.94, the reliability of the total score is quite high, but not unexpected for a scale with more than 30 items.

Exhibit 6: Reliability of Scores within Each Domain of the ERA-38

Domain/subscale	Number of items	Alpha Coefficient
General Health	5	0.80
Cognitive Function	4	0.80
Mental Health	12	0.89
Functional Independence	5	0.77
Sexual Function	2	0.83
Pain	2	0.58
Sleep	2	0.73
Fatigue	4	0.79
Urinary Incontinence	1	-
Appearance	1	-

[4] Note that alpha coefficients could not be computed for the two domains represented by a single item.

Domain/subscale	Number of items	Alpha Coefficient
Total Score	*38*	*0.94*

To examine the concurrent and discriminant validity of the instrument, scores on the various subscales were correlated with: (1) scores on the Mental (MCS-12) and Physical (PCS-12) components of the Medical Outcomes Study Short Form-12 (SF-12), (2) scores on a 13-item scale measuring the ability to carry out activities of daily living (ADL), (3) the number of comorbidities reported on the Charlson Comorbidity Scale, and (4) scores derived from a 5-item Geriatric Depression Scale (GDS). As expected, scores on the Mental Health subscale correlated most highly with scores on the MCS-12 (0.40) and scores on the GDS (-0.40). Also as expected, scores on the Functional Independence subscale correlated most highly with scores on the PCS-12 (0.32). The correlation between the ERA-38 total score and the number of comorbidities (-0.09), however, was much lower than expected. Although the authors concluded that these results generally support the concurrent validity of the tool, it is important to note that scores on the General Health, Mental Health, and Functional Independence subscales displayed statistically significant correlations across the board with the MCS-12, PCS-12, ADL, and GDS scales.

In subsequent work, Sarkisian, Steers, Berry, and Mangione (2005) developed a 12-item version of the instrument covering two of the original 10 domains in the ERA-18—Mental Health and Cognitive Function and one domain—Physical Health, consisting of items from three of the original domains. To develop the scale, the researchers conducted an exploratory factor analysis with data from the Sarkisian et al. (2002) study. The analysis revealed that the ERA-38 measures one dominant factor accounting for close to 37 percent of the variance in the scores, and six minor factors, each accounting for a small percent of the total variance in the scores. Based on the results of the factor analysis and the domains most highly valued by the focus group participants in the 2002 study, Sarkisian et al. (2005) selected three of the seven factors—Mental Health, Cognitive Function, and Physical Health—and four items from each of those factors according to the strength of their factor loadings for the 12-item version of the ERA (ERA-12). Items selected for the Mental Health and Cognitive Function subscales appear in those domains in the ERA-38. Items selected for the Physical Health subscale, on the other hand, appear in three different domains in the ERA-38. Two were from the ERA-38 General Health domain, one from the ERA-38 Pain domain, and one from the ERA-38 Fatigue domain, all of which exhibited strong loadings on the Physical Health factor uncovered in the exploratory factor analysis.

To validate the factor structure of the shortened instrument, Sarkisian et al. (2005) conducted a confirmatory factor analysis with data collected from an independent sample of 643 older adults. The three-factor model fit the data quite well. The researchers also examined the reliabilities of the ERA-12 total score and subscores in the samples from the 2002 and 2005 studies. Exhibit 7 displays the results from that examination, along with test-retest reliabilities obtained in a subsample of participants in the 2005 study who completed the ERA-12 a second time two weeks after they completed the tool the first time.

Exhibit 7: Reliabilities of the ERA-12 Total and Subscale Scores

| Domain/subscale | Number of items | Alpha Coefficient | | Test-Retest Reliability 2005 Subsample |
		2002 Study Sample	2005 Study Sample	
Mental Health	4	0.75	0.76	0.83
Cognitive Function	4	0.81	0.76	0.81
Physical Health	4	0.79	0.80	0.78
Total Score	*12*	*0.88*	*0.89*	*0.94*

In sum, the ERA-38 was developed without the benefit of a factor analysis to inform the structure and dimensionality of the tool. Although the subscales in the ERA-38 exhibited favorable properties for the most part, the results of a factor analysis of the data from the 2002 study in a subsequent study strongly suggests that the ERA-38 measures fewer than 10 dimensions or domains and that several items in the ERA-38 are not placed in the domains/dimensions they best represent. *Consequently, the extent to which the ERA-38 subscales measure distinct dimensions is not clear, especially when one considers that scores on several of the subscales displayed statistically significant correlations with other measures intended to validate the Mental Health and Functional Independence subscales. For that reason, we advise against the use of the ERA-38 subscales.*[5] The structure of ERA-12, on the other hand, was evaluated twice, once through an exploratory analysis of data from the 2002 study and once through a confirmatory analysis of data from the 2005 study, yielding similar results and supporting the replicability of the structure across independent samples of older adults. Across both samples, the ERA-12 subscales measure expectations regarding aging in three areas with acceptable to good reliability and expectations regarding aging in general with good reliability nearly matching the performance of the ERA-38. Moreover, according to Sarkisian et al. (2005), the ERA-12 captures 88 percent of the variance in the ERA-38. *These results suggest that the ERA-12 should be the instrument of choice, especially when administration time is a matter of concern since the majority of older adults are able to complete the tool within 5 minutes while the ERA-38 may take up to 15 minutes to complete*.

Image of Aging Scale

Levy and colleagues built on previous measures by noting that, while there were scales that measure knowledge of aging and negative expectations about aging, there was no scale to assess both positive and negative perceptions that individuals hold about older people (Levy, Kasl, & Gill, 2004). As a corrective, they developed the Image of Aging Scale. The Image of Aging Scale consists of 9 positive and 9 negative words or short phrases representing nine conceptual categories of perceptions individuals hold of old people (see Exhibit 8). Respondents are asked to report a "number from 0 to 6 that best shows how well

---

[5] One way to address this issue would be to examine the factor structure of the ERA-38 in the data collected if the size of the sample is sufficient.

[each] word or [phrase] matches [their] image or picture of old people in general, with 0 being furthest from what [they] think and 6 being the closest to what [they] think" (Levy et al., 2004, p. 2019).

Exhibit 8: Items in the Image of Aging Scale

Conceptual Category	Negative Attribute	Positive Attribute
Activity	walks slowly	active
Appearance	wrinkled	well-groomed
Cognition	senile	wise
Death	dying	full of life
Dependence	helpless	capable
Personality	grumpy	positive outlook
Physical Health	sick	healthy
Relationships	lonely	family-oriented
Will to Live	given up	will to live

Although the authors argue that the tool is suitable for measuring positive and negative perceptions of the old among individuals of all ages, to our knowledge the reliability and validity of the tool have only been examined in relatively small samples of older adults. In one study with 20 older adults with a mean age of 66, the positive subscale displayed a one-week test-retest reliability of 0.92, the negative scale, a one-week test-retest reliability of 0.79. In a larger sample of 68 older adults with a mean age of 69, the positive subscale yielded an alpha coefficient of 0.84, the negative scale, a coefficient of 0.82. With regard to its validity, Levy et al. (2004) reported evidence suggesting that scores on the negative subscale are positively related to lifetime exposure to television. *Although the results from the preliminary studies of the tool are promising, additional work is needed to thoroughly evaluate the tool's measurement properties and to assess whether the tool is suitable for measuring ageism with good reliability in younger adults.*

Ambivalent Ageism Scale (AAS)

Cary and his collaborators noted that there are a plethora of existent instruments designed to measure hostility towards older adults, but none that take benevolent (or patronizing) ageism into account even though the latter may be more widespread given its greater degree of social acceptability. In response, they developed the Ambivalent Ageism Scale, a 13-item survey meant to detect both benevolent and hostile ageism (Cary, Chasteen, & Remedios, 2017). The AAS tool asks respondents to indicate their level of agreement or disagreement with each of 13 statements (9 benevolent and 4 hostile) on a seven-point scale ranging from strongly disagree to strongly agree. Exhibit 9 displays items representing each type of statement.

Exhibit 9: Select Items from the Ambivalent Ageism Scale (AAS)

Type of statement	Sample Statements
Benevolent	Even if they want to, old people shouldn't be allowed to work because they are fragile and may get sick.
	Older people need to be protected from the harsh realities of our present society.
	It's helpful to repeat things to old people because they rarely understand the first time.
Hostile	Old people are easily offended.
	Old people exaggerate the problems they have at work.
	Old people are a drain on the health care system and the economy.

The AAS was developed and validated in four stages. In the first stage, 397 undergraduate students responded to 41 draft items designed to measure benevolent attitudes. Based on the information collected, Items with low average scores or poor face validity were eliminated from the pool, as were items that seemed redundant with other items in the pool. In the second stage, the 21 remaining items were administered to 194 participants, ages 18 to 34, recruited from Amazon's Mechanical Turk (MTurk). An exploratory factor analysis performed on the data yielded three factors with eigenvalues above 1.0. The first factor was defined by items related to cognitive assistance or physical protection, the second factor by two items related to asking for help, and the third factor by three items stating that it is good for old people to live in nursing homes. Because the third factor appeared to measure attitudes related to segregating old people from others rather than benevolent attitudes associated with vulnerability as intended, items loading on that factor were eliminated from the pool along with items exhibiting comparatively low factor loadings on the first two factors, leaving nine benevolent items in all.

In the third stage, four items representing hostile attitudes were added to the 9 benevolent items to form the final version of the AAS. The AAS, along with the Fraboni Scale of Ageism, the Warmth and Competence Stereotypes scale[6], and a demographic background questionnaire, were administered to 161 participants, age 18 to 57, recruited from MTurk. To examine the impact of adding the items measuring hostile items to the tool, two- and three-factor models were fit to the data. In the three-factor exploratory analysis, the benevolent items defined two of the three factors; the hostile items defined the other factor. However, the pattern of loadings across the two benevolent factors differed somewhat from the pattern found in the factor analysis performed in the second stage of the tool's development. In the two-factor solution, items measuring benevolent attitudes defined one factor; the four items measuring hostile attitudes defined the other factor. Of note, two of the 9 benevolent items exhibited loadings below 0.40 on that factor, a criterion often used when deciding whether to keep an item in a scale. Nonetheless, the scores on the 9-item benevolent scale exhibited good reliability (0.89), as did the scores on the 4-item hostile scale (0.84), and the scores on all 13 items in AAS (0.91).

---

[6] The Warmth and Competence Scale asks respondents to rate the extent to which a series of adjectives (e.g., sociable, kind, able, skillful) are descriptive of older adults on a 9-point scale.

To assess the convergent and discriminant validity of the AAS, the researchers correlated the total and subscale scores on the AAS with the total scores on the 29-item Fraboni Scale of Ageism. The AAS total scores were highly correlated (0.65) with the Fraboni scores. Given that more than 75 percent of the items in the Fraboni scale are hostile items, it came as no surprise that the AAS hostile subscale scores correlated more highly with the Fraboni scores (0.75) than the AAS benevolent subscale scores did (0.51). As expected, higher scores on the hostile subscale were predictive of lower scores on the Warmth and Competence subscales, and higher scores on the benevolent subscale were predictive of higher scores on the Warmth subscale. In the fourth and final stage of development, a small group (N=23) of participants were asked to complete the AAS two weeks after they completed it the first time to obtain estimates of test-retest reliability. The test-retest reliabilities for the total, hostile, and benevolent subscales were 0.80, 0.76, and 0.76, respectively.

In sum, the research conducted to date provides some evidence that the AAS measures two types of ageist attitudes, benevolent and hostile attitudes. Evidence supporting the validity of the hostile attitudes scale is stronger than that for the benevolent scale, even though the hostile scale consists of fewer items and is less reliable than the benevolent scale. The tool would benefit from further research to confirm its factor structure and validate its subscales. Of note, the relative size of the eigenvalues generated in the factor analyses suggests that the tool measures one dominant factor. The reliability of the total scale (0.91) suggests that it does so with good reliability.

II. Measuring Prejudice and Affect

Although early studies on ageism focused on whether younger people held stereotypes about the elderly or the aging process, researchers soon began to focus more specifically on the affective component of ageism as a better indicator of age bias or prejudice.

Fraboni Scale of Ageism (FSA)

The Fraboni Scale of Ageism (FSA) was developed to measure affective components of ageism (Fraboni, Salstone, & Hughes, 1990). These authors argued that previous research only addressed stereotyping and misconceptions surrounding aging, while neglecting the attitudinal facet of ageism. The tool asks respondents to indicate their level of agreement with each of 29 statements on a 4-point scale from strongly disagree to strongly agree. Statements representing each proposed dimension are shown in Exhibit 10.

Exhibit 10: Select Items from the Frabroni Scale of Ageism

Dimension/ Subscale	Sample Statements
Antilocution	Many old people are stingy and hoard their money and possessions.
	Most old people should not be trusted to take care of infants.

	Old people complain more than other people do.
Avoidance	I sometimes avoid eye contact with old people when I see them.
	Old people should find friends their own age.
	I don't like it when old people try to make conversation with me.
Discrimination	The company of most old people is quite enjoyable.
	It is best that old people live where they won't bother anyone.
	Old people don't really need to use our community sports facility.

To validate the measure, Fraboni et al. (1990) performed an exploratory factor analysis based on data from 100 high school students. The analysis suggested that the tool measures three dimensions of ageism: Antilocution (antagonism and antipathy towards older individuals), Avoidance (avoiding social contact with older individuals), and Discrimination (holding discriminatory beliefs regarding the rights and activities of older individuals).   The Antilocution, Avoidance, and Discrimination factors accounted for 23.3 percent, 7.2 percent, and 7.0 percent of the variance in the scores, respectively. The corresponding subscales yielded reliabilities of 0.76, 0.77, and 0.65 in the Fraboni et al.(1990) study of high schoolers, and reliabilities of 0.75, 0.61 and 0.77 in a subsequent study of college students (Rupp, Vodanovich, & Crede, 2005). Fraboni et al. (1990) also found that the responses across the 29 items were internally consistent with an alpha coefficient of 0.86. North and Fiske (2013) obtained comparable estimates of reliability for the total scores (0.86 and 0.91) in two samples of adults age 18 to 59 and age 18 to 60. Similarly, Cary, Chasteen, and Remedios (2017) reported a reliability coefficient of 0.93 in a sample of adults age 18 to 57.

In a subsequent study of the factor structure of the FSA, Rupp et al. (2005) administered the tool to a sample of 353 college students ranging in age from 17 to 58. The vast majority of participants (86.3 percent) in their study were under the age of 30, 7.55 percent were age 30 to 39, and 6.15 percent, age 40 or older. Close to 80 percent of the participants were white. Rupp and his coauthors (2005) found that a confirmatory factor analysis of the Fraboni three-factor structure did not fit their data well. To further examine the dimensional structure of the tool, the authors performed an exploratory factor analysis on data from a smaller sample of 201 undergraduates with background characteristics similar to those in the larger sample. The analysis yielded a three-factor solution that differed somewhat from that of Fraboni et al. (1990) with regard to the items loading on each factor. In addition, six of the items in the original FSA scale did load well on any of three factors.

On the basis of the structure uncovered in the exploratory factor analysis, Rupp et al. (2005) renamed the factors Stereotypes, Separation, and Affective attitudes. The Stereotypes factor was largely defined by items loading on Fraboni et al.'s (1990) Antilocution factor; the Separation factor, by items loading on the Avoidance factor and a few items loading on the Antilocution and Discrimination factors; and the Affective Attitudes factor, by items loading on the Discrimination factor and one item loading on the Avoidance factor. To validate the revised structure, a three-factor confirmatory factor model was fit to the data from the original sample after removing the six poorly performing items. A one-factor model was also fit to the data. Although the authors claimed that the revised three-factor model fit the data from the

23-item version of the FSA better than Fraboni's three-factor model fit the data from the original 29-item version of the FSA, the fit statistics from both models were similar in value and neither model provided what would be considered a "good" fit to the data. Moreover, the percent of variance in the scores accounted for by the revised model (36.4 percent) was similar in value to that accounted by Fraboni's model (37.5 percent). The Stereotypes, Separation, and Affective Attitudes subscales did, however, yield alpha coefficients that were higher than those found for two of the three corresponding Fraboni subscales (0.79 v. 0.75, 0.76 v. 0.61, and 0.70 v. 0.77, respectively). Unfortunately, the authors did not provide an estimate of the alpha coefficient for the 23-item scale.

To assess the concurrent validity of the revised FSA, Rupp et al. (2005) correlated the four sets of factor scores from the FSA with the scores from the subscales of the Aging Semantic Differential scale (ASD) and the scores from the OP- (negative) subscale of Kogan's Attitudes toward Old People Survey (OP). Of the four sets of factor scores, the Stereotypes scores correlated most highly with the subscale scores of the other ageism tools. Rupp et al. (2005) argued that the finding was expected since the content of the Stereotypes subscale is more closely aligned with the content of the other ageism scales. *In sum, the FSA is likely to measure more than one dimension of ageism. The tool would benefit from further research to clarify its dimensional structure and to improve upon the measurement properties of its subscales. Although research conducted to date does not support the use of the subscale scores, it does suggests that the total score on the 29-item version of the tool can be used to measure ageism with good reliability in high school students and in adults age 18 or older.*

The Succession, Identity, and Consumption (SIC) Ageism Scale

Conceived as an alternative to social identity theory, the intergroup-tension model of ageism theorizes that the elderly are in competition with younger adults for resources in 3 categories (succession, consumption, and identity), which leads to ageist bias among younger people who resent older adults for staying in the workforce longer, taking more than their "fair share" of public resources, and appropriating elements of the youth culture. The Succession, Identity, and Consumption (SIC) ageism scale was developed by North and Fiske (2013) to provide a measure that aligns with this theory of ageism. It differs from other instruments in that it focuses on prescriptive, or hostile, stereotypes. It is designed to capture beliefs about what older individuals should and should not do with regard to Succession (stepping aside from their positions and areas of influence), Identity (participating in activities typically reserved for younger individuals), and Consumption (consuming shared resources, such as healthcare) (North & Fiske, 2013). The tool asks respondents to indicate their level of agreement with each of 20 statements on a six-point scale from strongly disagree to strongly agree. Statements representing the three dimensions are shown below in Exhibit 11.

Exhibit 11: Select Items from the Succession, Identity, and Consumption (SIC) Scale

Dimension/ Subscale	Sample Statements
Succession	Most older people don't know when to make way for younger people.

Dimension/ Subscale	Sample Statements
	The older generation has an unfair amount of political power compared with younger people.
	Younger people are usually more productive than older people at their jobs.
Identity	Older people typically shouldn't go to places where younger people hang out.
	Older people probably shouldn't use Facebook.
	Generally older people shouldn't go clubbing.
Consumption	Older people are too big a burden on the health care system.
	Older people don't really need to get the best seats on buses and trains.
	AARP (American Association of Retired Persons) wastes charity money.

To validate the tool, the authors conducted four studies in which they examined the factor structure of the tool, the correlation of the SCI scores with other measures, and the internal consistency/reliability (Cronbach's alpha) of the SIC total score and the SIC subscores. Participants for the studies were recruited online from Princeton University and from Amazon Mechanical Turk. The samples ranged in size from 93 to 1,283 and in age from 16 to 81, 18 to 60, 18 to 81, and 18 to 59 across the four studies. About three-fourths of the participants in each sample were white. Across all four studies, a three-factor confirmatory model fit the data well and provided a better fit than a one-factor model. Despite differences in the size and composition of the samples, the results of the factor analyses were remarkably similar across the studies as were the estimates of reliability for the SCI total score (0.90 to 0.91). The subscale scores tended to be less reliable than the total score as would be expected given that the subscores are based on fewer items than the total score. The reliabilities of the subscores were also more variable than those for the total score, ranging in value from 0.84 to 0.85 on the Succession subscale, from 0.83 to 0.87 on the Identity subscale, and from 0.75 to 0.86 on the Consumption subscale.[7]

Of the other measures of prejudice included in the study (i.e., the Ambivalent Sexism Inventory, the Symbolic Racism Scale, and the Fraboni Scale of Ageism), the SCI total score correlated most strongly with the Fraboni Scale of Ageism (0.70 v. 0.32-0.40) as expected. Also as expected, younger individuals and men tended to score higher on the scale and on the three subscales than older individuals and women. *Overall, the SCI has consistently measured prescriptive ageism with good reliability in samples covering a wide range of age. The research conducted to date supports the use of the total score as a reliable and valid indicator of overall prescriptive ageism and the use of subscores as reliable and valid measures of different aspects of prescriptive ageism.*

Pigram Ageism Scale

Braithwaite, Lynd-Stevenson, and Pigram (1993) developed the Ageism Scale using work from Pigram's thesis in order to address their concerns that definitions of "ageism" offered up until this point were

---

[7] Note that reliabilities of 0.90 or above are typically regarded as excellent, those between 0.80 and 0.89 as good, those between 0.70 and 0.79 as acceptable, and those below 0.70 as poor or unacceptable.

difficult to parse, and therefore it was difficult to legitimately study the concept and the relationships between its component parts: stereotype, attitude, and behavior. The scale focuses on stereotypes and attitudes, setting aside behavior, although it does include some behaviorally-relevant measures, such as contact with elderly people and age-based expectations about employment and hiring processes.

Specifically, Pigram developed five scales of ageism designed to measure: 1) attitudes toward the elderly, 2) attitudes toward the aging process, 3) discrimination against the elderly in society 4) stereotypes regarding the capability of the elderly, and 5) stereotypes regarding the sociability of the elderly. Each scale consists of an equal number of positively and negatively worded statements. The first three scales ask respondents to indicate the extent to which they agree or disagree with each statement on a six-point scale from strongly agree to strongly disagree. The last two scales ask respondents to indicate the extent to which they believe each statement is true or false on a five-point scale from always true to always false. Exhibit 12 displays two statements from each of the five scales, along with the number of items in each scale.

Exhibit 12: Select Statements from the Five Pigram Ageism Scale

Scale	Number of items	Sample Statements
Attitudes toward the elderly	8	I really enjoy talking to older people.
		Older people really don't say much or do much that is of interest to me.
Attitudes toward the aging process	16	In my old age I will be as enthusiastic about life as I am now.
		Once you get to a certain age, life inevitably goes downhill.
Discrimination against the elderly	10	This is a youth oriented society.
		Old people in this country are treated with sympathy and understanding.
Stereotypes regarding capability	10	Older people lose the ability to pay attention to detail.
		Older people are quite capable of performing tasks that require effort and stamina.
Stereotypes regarding sociability	8	Older people are less understanding of other people's problems.
		Other people are more sympathetic in their dealings with other people.

To examine the properties of the scales and their relationship with other measures, Braithwaite, Lynd-Stevenson, and Derek (1993) administered the scales to 195 first-year psychology students, age 16 to 62, over the course of two years. The researchers also collected information on the sex and age of the participants, and the frequency with which they had contact with their grandparents or elderly friends as recorded on a seven-point scale ranging from several times a week to less than once a year. Exhibit 13 shows that all of the scales exhibited reliabilities of 0.80 or above, with the exception of the scale measuring stereotypes regarding the sociability of the elderly, which yielded an acceptable reliability of 0.76.

Exhibit 13: Reliability of the Ageism Scales and Correlations with other Measures

Scale	Reliability	Correlations		
		Frequency of Contact	Age	Sex
Attitudes toward the elderly	0.81	-0.37*	0.19*	-0.23*
Attitudes toward the aging process	0.86	-0.21*	0.22*	-0.03
Discrimination against the elderly	0.82	0.18*	-0.07	0.04
Stereotypes regarding capability	0.80	-0.12	0.15*	0.00
Stereotypes regarding sociability	0.76	-0.06	-0.02	-0.07

*statistically significant

With regard to the correlations of the scales with other measures, the exhibit reveals that older students tend to have more positive attitudes toward older people and the aging process than do younger students. Likewise, older students tend to believe that older people are more capable than do younger students. The exhibit also shows that females tend to express more positive attitudes toward the elderly than males. Finally, students who have more regular contact with the elderly tend to exhibit more positive attitudes toward the elderly and the aging process and tend to believe that society discriminates against the elderly than do students who have less regular contact with the elderly. *While the five scales performed well in this study, further studies are needed to fully assess the measurement properties of the scales*.

Implicit Association Test (IAT)

The Implicit Association Test (IAT) is an instrument meant to measure the strength of automatic associations between a target concept (e.g. age) and an attributive concept (e.g. good or bad). Developed in 1998, this somewhat controversial measure has been widely employed in social psychology research and proved to have high predictive validity in a variety of contexts (Greenwald, Banaji, & Nosek, 2015). The IAT was recently used to study implicit ageist stereotypes and attitudes (Busso, Volmert, & Kendall-Taylor, 2018). Busso et al sought to investigate the malleability of implicit age bias in a nationally-representative sample of American adults by presenting subjects (N=767) with messaging about aging, and subsequently administering the IAT. The randomly-assigned messages either: (a) framed aging in a positive light or discussed the negative effects of ageism, (b) presented unframed information about aging, or (c) presented unrelated information to a control group. The study found that those participants who received framing messaging about aging displayed lower levels of implicit bias, relative to those who received unframed or unrelated messaging, suggesting that implicit prejudice can be mitigated by positive messaging regarding aging. *However, the psychometric properties of the IAT in studying ageism are as yet untested*.

III. Measuring Discriminatory Behavior

While the stereotyping and prejudice components of ageism have received significant attention in the literature over the past 50 years, relatively less work has sought to identify and capture the discriminatory behavioral component – especially with regards to personal-level discrimination. Although not a focus of this review, we provide an example each of tools that attempt to measure the extent to which people engage in or experience age discrimination since these could in principle be used to study the impact of an intervention designed to reduce ageist stereotypes.

The Ageism Survey

Palmore (2001) developed the Ageism Survey to measure the extent to which older adults experience various types of ageism. The tool asks respondents to indicate how often they have experienced each of 20 events on a three-point scale from 0 to 2, where 0 indicates never, 1 indicates once, and 2 indicates more than once. Exhibit 14 shows a few items in the tool.

Exhibit 14: Select Items from the Ageism Survey

I was told a joke that pokes fun at older people.
A doctor or nurse assumed my ailments were caused by my age.
Someone assumed that I could not hear well because of my age.

The psychometric properties of the tool were examined in a convenience sample of 84 older individuals, age 60 to 93 who varied in their levels of educational attainment. Thirty percent of the sample had completed at least a bachelor's degree, 31 percent, at least some college. The other 39 percent either had earned a high school diploma or did not finish high school. A principal components factor analysis performed on the data yielded an eigenvalue of 4.74, suggesting that a single factor accounts for 24 percent of the variance in the scores and leading the author to conclude that the instrument measured one dominant dimension. An alpha coefficient of 0.81 was obtained when all 20 items were included in the calculation. The author noted that three of the items were rarely endorsed (presumably meaning that most respondents indicated that they had never experienced the event). Even though eliminating those items from the scale raised the value of the alpha coefficient, Palmore (2001) chose to keep them in since they represent three of the most serious types of ageism—discrimination in employment, leadership, and rental housing. *These results suggest that the Ageism Survey is a promising tool that exhibits good reliability in the population it is intended to assess. Although it makes sense substantively to retain the three rarely endorsed items in the instrument, responses to those items are most likely adding error variance to the scores. When using the scale, it would be best to score the instrument with and without those items to assess whether keeping them in affects the conclusions drawn.*

The Relating to Older People Evaluation (ROPE)

Cherry and Palmore (2008) developed the Relating to Older People Evaluation (ROPE) survey to detect engagement in positive and negative ageist behaviors in everyday life and to help assess: the prevalence of ageist behavior in society, the relative prevalence of specific types of ageist behavior, and what types of people report more or less ageist behavior. The ROPE is a 20-item self-report questionnaire that asks respondents to indicate how frequently (never, sometimes, or often) they engage in six positive and 14 negative types of ageist behaviors. Statements representing both types of behavior are shown in Exhibit 15.

Exhibit 15: Select Items in the Relating to Older People Evaluation Scale

Dimension/ Subscale	Sample Statements
Positive Behavior	Hold doors open for old people because of their age.
	Vote for an old person because of their age.
Negative Behavior	Send birthday cards to old people that joke about their age.
	Vote against an old person because of their age

The researchers administered the instrument to a sample of 314 college students, community-dwelling older adults, and various members of a university community. However, the psychometric evaluation of the tool appears to have been limited to the analysis of data from the 90 college students who completed the tool twice. The analysis examined the test-retest reliability of the students' scores on the positive and negative items, and the reliability of their scores on the items as a whole, yielding values of 0.57, 0.72, and 0.70, respectively. *These results suggest that the ROPE requires further development to measure ageism behavior with good reliability. Further research is also needed to establish the tool's validity and its suitability for measuring ageism behavior among individuals of all ages.*

Conclusion: Where Do We Go from Here?

Our review suggests that not all tools used by researchers are created equal, with some measuring different (or unknown) psychological dimensions of ageism and others having questionable (or untested) psychometric properties. However, designing interventions to combat ageism among young people requires a strong evidence base, which in turn necessitates the use of more rigorous research methods than have typically been used. In her systematic review of research on such interventions, Chonody (2015) finds that very few meet basic standards for scientific research and only 35 of 58 included studies used a pre-/post-test design. Even fewer included a control group (n=12) or conducted any follow-up to determine if changes are lasting. Another common methodological weakness noted by Chonody is the lack of psychometric testing and reliance on measures such as Palmore's FAQ (n=12) that may not be appropriate for the samples used or even capture the underlying psychological construct of interest,

measuring factual knowledge as a proxy for holding ageist stereotypes. Other commonly used measures such as the Aging Semantic Differential (n=12), Fabroni's Scale of Ageism (n=2), and Kogan's Attitudes Towards Old People (n=5) measure attitudes or bias more directly but leave open the question of whether holding ageist stereotypes results in actual discrimination against the elderly.

Indeed, Voss and her coauthors (2018) review the literature on the relationship between psychological constructs used to study ageism but conclude that "it is much too early for sweeping conclusions regarding the influence of age stereotypes on age discrimination" (p. 25). One avenue for developing better theories is to consider different measures such as those used in recent studies "that have extended the boundaries of the stereotype-discrimination relations by using, for example, implicit stereotype measures, by considering additional variables, or by focusing on moderators" (ibid.). Fifty years after Butler's call to arms, much work remains to be done. Voss concludes her review of the field by tasking ageism researchers with multiple lines of investigation, ranging from theory development to more rigorous research designs to finer-grained hypothesis testing:

> Future research on the relations between age stereotypes and age discrimination is urgently needed that (a) is based on an adequate definition and assessment of age discrimination, (b) combines field studies (typically using a correlative design) with more controlled experiments in the lab, and (c) takes into consideration moderating variables in order to investigate for whom and in which situations the relation holds or does not hold.

References

Barron, M. L. (1953). Minority Group Characteristics of the Aged in American Society. *Journal of Gerontology, 8*(4), 477-482. DOI: https://doi.org/10.1093/geronj/8.4.477

Berger, R. (2017). Aging in America: Ageism and General Attitudes toward Growing Old and the Elderly. *Open Journal for Social Sciences, 5*(8), 183-198.

Blakeborough, D. (2008). "Old People Are Useless": Representations of Aging on *The Simpsons*. *Canadian Journal on Aging, 27*(1), 57-67. DOI: https://doi.org/10.3138/cja.27.1.57.

Braithwaite, V., Lynd-Stevenson, R., & Pigram, D. (1993). An empirical study of ageism: From polemics to scientific utility. *Australian Psychologist, 28*, 9-15.

Busso, D. S., Volmert, A., & Kendall-Taylor, N. (2018). Reframing Aging: Effect of a Short-Term Framing Intervention on Implicit Measures of Age Bias. *Journals of Gerontology: Psychological Sciences, 20*(20), 1-6.

Butler, R. N. (1969). Age-ism: Another Form of Bigotry. *The Gerontologist, 9*, 243-246.

Butler, R. N. (1980). Ageism: A Foreword. *Journal of Social Issues, 36*(2), 8-11.

Cary, L. A., Chasteen, A. L., & Remedios, J. (2017). The Ambivalent Ageism Scale: Developing and Validating a Scale to Measure Benevolent and Hostile Ageism. *The Gerontologist, 57*(2), e27-e36. DOI: https://doi-org.proxy.uchicago.edu/10.1093/geront/gnw118

Cherry, K. E., Allen, P. D., Denver, J. Y., & Holland, K. R. (2015). Contributions of Social Desirability to Self-Reported Ageism. *Journal of Applied Gerontology, 34*(6), 712-733. DOI: 10.1177/0733464813484984

Cherry, K. E., & Palmore, E. (2008). Relating to Older People Evaluation (ROPE): A Measure of Self-Reported Ageism. *Educational Gerontology, 34*(10), 849-861. DOI: 10.1080/03601270802042099.

Chonody, J. M. (2015). Addressing Ageism in Students: A Systematic Review of the Pedagogical Intervention Literature. *Educational Gerontology, 41*, 859-887. DOI: 10.1080/03601277.2015.1059139.

Couto, P., Pinheiro, M. C., & Wentura, D. (2012). Automatically activated facets of ageism: MAsked evaluative priming allows for a differentiation of age-related prejudice. *European Journal of Social Psychology, 42*(7), 852-863.

Cowan, D. T., Fitzpatrick, J. M., Roberts, J. D., & While, A. E. (2004). Measuring the Knowledge and Attitudes of Health Care Staff Toward Older People: Sensitivity of Measurement Instruments. *Educational Gerontology, 30*, 237-254.

Drake, J. T. (1959). The Aged in American Society. *Science and Society, 23*(2), 176-177.

Fraboni, M., Saltstone, R., & Hughes, S. (1990). The Fraboni Scale of Ageism (FSA): An Attempt at a More Precise Measure of Ageism. *Canadian Journal on Aging / La Revue Canadienne Du Vieillissement, 9*(1), 56-66. DOI: 10.1017/S0714980800016093.

Fung, H. H., Li, T., Zhang, X., Sit, I. M. I., Cheng, S.-T., & Isaacowitz, D. M. (2015). Positive Portrayals of Old Age Do Not Always Have Positive Consequences. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 70*(6), 913-924. DOI: 10.1093/geronb/gbu061.

Gonzales, E., Tan, J., & Morrow-Howell, N. (2010). Assessment of the Refined Aging Semantic Differential: Recommendations for Enhancing Validity. *Journal of Gerontological Social Work, 53*(4), 304-318. DOI: 10.1080/01634371003715791

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically Small Effects of the Implicit Association Test Can Have Societally Large Effects. *Journal of Personality and Social Psychology, 108*(4), 553-561.

Harwood, J. (2007). Mass Communication Portrayals of Older Adults *Understanding Communication and Aging: Developing Knowledge and Awareness.* London: Sage Publications, Inc.

Hilt, M. L., & Lipschultz, J. H. (1999). Revising the Kogan Scale: a test of local television news producers' attitudes toward older adults. *Educational Gerontology, 25*(2), 143-153.

Iwasaki, M., & Jones, J. A. (2008). Attitudes Toward Older Adults: A Reexamination of Two Major Scales. *Gerontology & Geriatrics Education, 29*(2), 139-157. DOI: 10.1080/02701960802223209

Kite, M. E., & Johnson, B. T. (1988). Attitudes Toward Older and Younger Adults: A Meta-Analysis. *Psychology and Aging, 3*(3), 233-244.

Kite, M. E., Stockdale, G. E., Whitley, J., Bernard E., & Johnson, B. T. (2005). Attitudes Toward Younger and Older Adults: An Updated Meta-Analytic Review. *Journal of Social Issues, 61*(2), 241-266.

Kogan, N. (1961). Attitudes toward old people: The development of a scale and an examination of correlates. *Journal of Abnormal Psychology, 62*(1), 44-54.

Laidlaw, K., Power, M. K., Schmidt, S., & WHOQOL-OLD Group. (2007). The attitudes to ageing questionnaire (AAQ): development and psychometric properties. *International Journal of Geriatric Psychiatry, 22*, 367-379. DOI: 10.1002/gps.1683

Lasher, K. P., & Faulkender, P. J. (1993). Measurement of Aging Anxiety: Development of the Anxiety about Aging Scale. *The International Journal of Aging and Human Development, 37*(4), 247-259. DOI: https://doi-org.proxy.uchicago.edu/10.2190/1U69-9AU2-V6LH-9Y1L

Lauzen, M. M., & Dozier, D. (2005). Maintaining the Double Standard: Portrayals of Age and Gender in Popular Films. *Sex Roles: A Journal of Research, 52*(7), 437-446. DOI: 10.1007/s11199-005-3710-1

Levy, B. (2009). Stereotype Embodiment: A Psychosocial Approach to Aging. *Current Directions is Psychological Science, 18*(6), 332-336.

Levy, B. R., Chung, P. H., Bedford, T., & Navrazhina, K. (2013). Facebook as a Site for Negative Age Stereotypes. *The Gerontologist, 52*(2), 172-176.

Levy, B. R., Kasl, S. V., & Gill, T. M. (2004). Image of Aging Scale. *Perceptual and Motor Skills, 99*, 208-210.

Levy, S. R., & Macdonald, J. L. (2016). Progress on Understanding Ageism. *Journal of Social Issues, 72*(1), 5-25.

Loos, E., & Ivan, L. (2018). Visual Ageism in the Media *Contemporary Perspectives on Aging* (pp. 163-176): Springer.

Miller, D. W., Leyell, T. S., & Mazachek, J. (2004). Stereotypes of the Elderly in U.S. Television Commercials from the 1950s to the 1990s. *The International Journal of Aging and Human Development, 58*(4), 315-340.

North, M. S., & Fiske, S. T. (2013). A prescriptive intergenerational-tension ageism scale: Succession, identity, and consumption (SIC). *Psychological Assessment, 25*(3), 706-713.

Osgood, C. E. & Suci, G. J. (1955). Factor Analysis of Meaning. *Journal of Experimental Psychology, 50*(5), 325-338.

Palmore, E. (1977). Facts on Aging: A Short Quiz. *The Gerontologist, 17*(4), 315-320. DOI: https://doi.org/10.1093/geront/17.4.315

Palmore, E. (1998). *The facts on aging quiz (2nd edition)*. New York: Springer.

Palmore, E. (2001). The Ageism Survey: First Findings. *The Gerontologist, 41*(5), 572-575. DOI: https://doi-org.proxy.uchicago.edu/10.1093/geront/41.5.572

Palmore, E. E., Branch, L. E., & Harris, D. E. (2005). *Encyclopedia of Ageism*. New York: Routledge.

Polizzi, K. G. (2003). Addressing Attitudes Toward the Elderly: Polizzi's Refined Version of the Aging Semantic Differential. *Educational Gerontology, 29*(3), 197-216.

Polizzi, K. G., & Steitz, J. A. (1998). Examining the Aging semantic differential: Suggestions for refinement. *Educational Gerontology, 24*(3), 207-223.

Popham, L. E., Kennison, S. M., & Bradley, K. I. (2011). Ageism, Sensation-Seeking, and Risk-Taking Behavior in Young Adults. *Current Psychology, 30*, 184-193. DOI: 10.1007/s12144-011-9107-0

Robinson, J. D., & Skill, T. (1995). The Invisible Generation: Portrayals of the Elderly on Prime-time Television. *Communication Reports, 8*(2).

Robinson, T., Gustafson, B., & Popovich, M. (2008). Perceptions of Negative Stereotypes of Older People in Magazine Advertisements: Comparing the Perceptions of Older Adults and College Students. *Ageing and Society, 28*, 233-251. DOI: https://doi.org/10.1017/S0144686X07006605

Rosencranz, H. A., & McNevin, T. E. (1969). A factor analysis of attitudes toward the aged. *The Gerontologist, 9*(1), 55-59. DOI: http://dx.doi.org/10.1093/geront/9.1.55

Rupp, D. E., Vodanovich, S. J., & Crede, M. (2005). The Multidimensional Nature of Ageism: Construct Validity and Group Differences. *The Journal of Social Psychology, 145*(3), 335-362. DOI: 10.3200/SOCP.145.3.335-362.

Sarkisian, C. A., Hays, R. D., Berry, S., & Mangione, C. M. (2002). Development, Reliability, and Validity of the Expectations Regarding Aging (ERA-38) Survey. *The Gerontologist, 42*(4), 534-542. DOI: https://doi-org.proxy.uchicago.edu/10.1093/geront/42.4.534

Sarkisian, C. A., Steers, W. N., Hays, R. D., & Mangione, C. M. (2005). Development of the 12-Item Expectations Regarding Aging Survey. *The Gerontologist, 45*(2), 240-248. DOI: https://doi-org.proxy.uchicago.edu/10.1093/geront/45.2.240

Smith, S. L., Pieper, K., & Choueiti, M. (2018). Still Rare, Still Ridiculed: Portrayals of Senior Characters On Screen in Popular Films from 2015 and 2016 *Media, Diversity, & Social Change Initiative, USC Annenberg School for Communication and Journalism*: Report prepared for Humana.

Tuckman, J., & Lorge, I. (1953). Attitudes Toward Old People. *The Journal of Social Psychology, 37*, 249-260.

Vasil, L., & Wass, H. (1993). Portrayal of the Elderly in the Media: A Literature Review and Implications for Educational Gerontologists. *Educational Gerontology, 19*(1), 71-85.

Voss, P., Bodner, E., & Rothermund, K. (2018). Ageism: The Relationship between Age Stereotypes and Age Discrimination *Contemporary Perspectives on Aging* (pp. 11-31): Springer.