

Article

Develop and Validate a Survey to Assess Adult's Perspectives on Autonomous Ridesharing and Ridehailing Services

Justin Mason ^{1,2,*}  and Sherrilene Classen ¹

¹ Department of Occupational Therapy, College of Public Health & Health Professions, University of Florida, Gainesville, FL 32603, USA; sclassen@phhp.ufl.edu

² Driving Safety Research Institute, College of Engineering, University of Iowa, Iowa, IA 52241, USA

* Correspondence: justin-mason@uiowa.edu

Abstract: Autonomous vehicles (AVs) have generated excitement for the future of transportation. Public transit agencies and companies (i.e., Uber) have begun developing shared autonomous transportation services. Most AV surveys focus on public opinion of perceived benefits and concerns of AVs but are not directly tied to field implementation of AVs. Experience and exposure to new technology affect adults' perceptions and level of technology acceptance. As such, the Autonomous RideShare Services Survey (ARSSS) was developed to assess adults' perceptions of AVs before and after being exposed to AVs. Face validity and content validity were established via focus groups and subject-matter experts (CVI = 0.95). Adults in the U.S. (N = 553) completed the ARSSS, and a subsample (N = 100) completed the survey again after two weeks. An exploratory and confirmatory factor analysis demonstrated that the ARSSS consists of three factors that can be used to reliably quantify users' perceptions of AVs: (a) Intention to Use, Trust, and Safety ($r = 0.85$, $p < 0.001$, ICC = 0.99); (b) Potential Benefits ($r = 0.70$, $p < 0.001$, ICC = 0.97); and (c) Accessibility ($r = 0.78$, $p < 0.001$, ICC = 0.96) of AVs. These are key factors in predicting intention to use and acceptance of AVs. Results from the ARSSS may inform the acceptance among users of these AV technologies.

Keywords: autonomous vehicles; ridesharing services; survey validation; perceptions of transportation; transportation habits



Citation: Mason, J.; Classen, S. Develop and Validate a Survey to Assess Adult's Perspectives on Autonomous Ridesharing and Ridehailing Services. *Future Transp.* **2023**, *3*, 726–738. <https://doi.org/10.3390/futuretransp3020042>

Academic Editors: Ouri E. Wolfson and Shunde Yin

Received: 31 March 2023

Revised: 4 May 2023

Accepted: 24 May 2023

Published: 1 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Automated vehicles have some level of driving automation to support or replace the driver. The Society of Automotive Engineers has defined levels of automation ranging from no automation (Level 0) to full automation (Level 5) [1]. The focus of this paper is on highly automated vehicles (Level 4) and fully automated vehicles (Level 5). Since the term 'automated' can refer to any level of automation, 'autonomous' will be used throughout the paper in reference to Levels 4 and 5. Autonomous vehicles (AVs) have safety, societal, and environmental implications and are expected to alter transportation systems. The potential benefits of AVs include improved road safety by mitigating crashes that occur due to human error, non-recurrent congestion impacts because of crash reduction, reduced pollution, and improved efficiency of transportation systems [2]. However, if society is to reap the benefits of AVs, the general public must accept AVs. Shared AVs will join similar shared mobility services such as car-sharing, bike-sharing, and demand-ride services. AVs may accelerate the growth of shared mobility services [3], and shared mobility services can make the deployment of AVs financially viable [4,5]. AV manufacturers and industrial partners are collaborating to develop and deploy AVs as shared autonomous mobility services. For example, there is significant interest and investment in shared autonomous transportation services from transportation network companies (i.e., Uber, Lyft, Beep). Despite the excitement, there is much uncertainty due to users' lack of trust, hesitation, and concerns about reliability pertaining to AVs. These concerns may decrease the likelihood that individuals include AVs in their transportation planning or daily commutes.

If policy makers, researchers, and manufacturers are to understand adults' perceptions of AVs, individuals should experience this technology to promote their familiarity with the current state of AVs—which may inform their perceptions pertaining to this technology. Although in the piloting phase, demonstration projects of shared AVs are occurring throughout the US [6] as well as in other nations [7]. Results from demonstration projects are promising regarding users' acceptance rates, perceived benefits, and intention to use AVs after experiencing autonomous shuttles or other vehicle types (e.g., Ford Transit, vans, cars) retrofitted with driving automation. Compared to older adults, younger adults tend to be more trusting of AVs and thus report a greater intention to use AVs [8]. Current findings in the literature comparing the perceptions of AVs between males and females are equivocal and tend to be more nuanced due to age classifications within sex (e.g., younger males vs. older females; [8]). While demographics such as age and sex provide some insight into predicting acceptance, additional factors should be considered, such as transportation habits, access to transportation, and (dis)ability status. Although several surveys have been developed to quantify user perceptions of AVs, most are focused on public opinion of AVs and are often abandoned prior to validation.

Researchers develop item pools that align with their research questions by modifying previously validated surveys or by generating items based on theoretical and conceptual underpinnings such as the Technology Acceptance Model, Unified Theory of Acceptance and Use of Technology, Car Technology Acceptance Model, Automated Vehicle User Perception Survey, and 4P Acceptance Model [9–13]. The use of an unvalidated survey can still provide insight [14–16] but does warrant concerns pertaining to the reliability and validity of the research findings. Alternatively, researchers use structural equation modeling (SEM) to enhance the validity and reliability of their survey results within their sample, but this does not support survey implementation and deployment for policymakers, transportation planners, or entities that want to use a validated survey to collect survey responses among a small sample or do not have the resources to perform exhaustive psychometric testing. Results from unvalidated surveys may be used in review papers [17], conflated in the news, used in a secondary data analysis [16], or detailed in articles geared toward the layperson. Since AVs are an emerging technology and are not yet widely deployed, it is important to have a valid and reliable tool that can provide scores to better understand trends and changes in users' perceptions of AVs and assess the effects of AVs on individuals' transportation habits.

Shared autonomous transportation services and AVs may result in reduced private car ownership or other massive changes to current transportation preferences or trends. To our knowledge, no survey has been constructed to reliably and validly measure adults' perceptions of autonomous ridesharing, ridehailing, and shuttles. Autonomous transportation services may also provide immediate and direct benefits for community mobility, especially among adults who are unable to drive, do not want to drive, or do not have adequate access to transportation (i.e., transportation disadvantaged). Individuals that are transportation disadvantaged often face barriers to participating in research and thus are often not represented in research. Projects that focus on transportation disadvantaged populations (e.g., impaired mobility, disabilities, driving cessation, older adults) are often underpowered or use surveys that are neither valid nor reliable. Developing and validating a survey is an important first step in understanding individuals' perceptions of AVs, transportation habits, and access to current modes of transportation. Such information can be used to better understand longitudinal trends or make comparisons across different populations in different geographic locations.

Therefore, the purpose of this study was to report on the item development, face, content, and construct validity, and 2-week test-retest reliability of a survey to quantify perceptions of older adults (>50 years of age) on AVs and capture transportation habits and demographics. Item development is the first phase of developing a survey and includes selecting items from previous surveys, reviewing the literature for important constructs, and generating items to overcome current gaps found in the literature. An

item pool must be evaluated, and items may be reduced by the research team or content experts to address redundant items, improve the flow of the survey, and reduce participant burden. Next, face validity can be established via focus groups that represent the intended audience of the survey. Face validity is an initial judgment of a survey's potential to assess the concepts it purports to measure and how items are interpreted by the intended audience [18]. Content validity is assessed to measure the extent to which a survey reflects a specific domain of content. Generally, face validity is more subjective than content validity and assesses whether the intended audience believes the items are suitable, sensible, appropriate, and relevant [19]. Content validity is assessed by subject-matter experts (SME) and determines if the items are fully representative of what the survey aims to measure. Lastly, construct validity assesses the quality of the relationship between items (and factors) but does not assess the extent to which a measure captures what it is intended to measure (i.e., content validity). During scale development, feedback from individuals that will administer or complete the survey can improve the acceptability, relevance, and quality of the measures [20].

After establishing face and content validity, the survey can then be deployed among a sample, and survey results can be psychometrically tested via factor analysis, Rasch modeling, Mokken scaling, or item response theory. During factor analysis, statistical techniques are used to reduce the dimensionality of the data and assess factor structure, item correlations, and if correlated items represent factors. Factor structure is essential in understanding, scoring, and interpreting survey responses [12,21]. Exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) are traditionally used to determine the optimal number of factors to retain in a model and can be used to inform and establish survey psychometrics. EFA is a suitable approach during the early stages of development and can inform the dimensionality of the survey by identifying relationships between measured variables [22]. A CFA can then be used to confirm the factor structure or theoretical model. While it is important to understand what a survey is measuring, it is crucial to have a reliable survey that will consistently produce similar results. Specifically, test-retest reliability is used to assess the consistency of test scores from one administration to the next. To assess test-retest reliability, the survey must be completed twice by the same respondents (i.e., single rater). Based on the purpose of the survey, tool, or assessment, further validation techniques can be deployed to reduce items to create a brief form of the survey or establish convergent, discriminant, or criterion validity.

2. Materials and Methods

This study was exempted by the University of Florida's Institutional Review Board (IRB201903309). All participants in the focus groups and on Amazon Mechanical Turk (MTurk) provided their written consent or waived consent to participate in the study. Participants were compensated 5.00 USD for completing the survey.

Team members reviewed both grey and scientific literature pertaining to driving and transportation habits (e.g., Driving Habits Questionnaire; [23]) and surveys examining adults' perceptions (perceived usefulness, perceived ease of use, intention to use, safety, trust, affordability, control and driving, accessibility, and social influences) of AVs, including the Technology Acceptance Model, Unified Theory of Acceptance and Use of Technology, Car Technology Acceptance Model, Automated Vehicle User Perception Survey, and 4P Acceptance Model [9–13]. The research team met to reduce the item pool with the goal of reducing redundancy, improving survey flow and clarity, and developing a survey that could be deployed before and after experiencing an autonomous ridesharing service. Feedback was solicited from stakeholders and the project sponsor prior to establishing face validity.

2.1. Face Validity and Content Validity

Face validity of the Autonomous RideShare Services Survey (ARSSS) was assessed by conducting two virtual focus groups with residents in two communities in Florida that

were vastly different based on socioeconomic status, access to transportation, and level of rurality (i.e., urban vs. rural) via a videotelephony software program (i.e., Zoom). Focus groups were conducted online in response to the pandemic to prevent unwarranted risk to the research team and older participants and promote research participation among the transportation disadvantaged. Each focus group consisted of five adults, a moderator (i.e., an expert in conducting focus groups for survey development), and a notetaker. Feedback from the focus groups was iteratively integrated, and each focus group met twice with the researchers. Participants in the focus groups provided feedback on the wording, meaning, clarity, credibility, and understandability of the items in the survey to remove jargon and promote comprehension at an eighth-grade reading level.

Prior to assessing content validity, Microsoft Word was used to assess the surveys' readability scores. The readability score (i.e., Flesch Reading Ease Score) was calculated based on the average number of syllables per word and the average sentence length. The Flesch Reading Ease Score rates text on a 100-point scale; the higher the score, the easier it is to understand the document. For most standard documents, the aim is a score of approximately 60 to 70. The Flesch–Kincaid Grade Level Score rates text on a U.S. grade-school level. For example, a score of 8 means that an eighth grader can understand the document. For most standard documents, the aim is for a score of approximately 7 to 8.

For content validity, three or more raters are required, but at least seven are suggested, to provide a content validity index, and raters with expertise in the content area under investigation (i.e., SMEs) are to be considered. Additional rounds of feedback from SMEs may be required to reach an acceptable level of agreement (i.e., content validity index > 0.90; [24]). To assess content validity, 10 SMEs were selected with broad but relevant expertise in rehabilitation science, traffic engineering, human factors, gerontology, psychology, transportation planning, and mobility as a service. The SMEs were sent content validity index (CVI) rater instructions and the ARSSS (i.e., 52 items) without the demographic items, as these items were developed with the sponsor (Florida Department of Transportation) to align with their previously constructed surveys. The SMEs provided feedback via a Qualtrics survey by rating the relevance of each item on a four-point Likert scale (1 = not relevant, 2 = relevant with major revisions, 3 = relevant with minor revisions, and 4 = very relevant). Feedback from SMEs was collated, and item-level CVI scores (i.e., the proportion of the ten raters who scored the item as relevant) and scale CVI scores were calculated [24]. Rater scores were collapsed, with an item-level score of 3 or 4 indicating acceptable item relevance and a score of 1 or 2 indicating the need for a major revision or low item relevance. Furthermore, SMEs provided feedback via open-ended responses to remove, refine, reword, or add survey items to enhance the content validity of the survey. The ARSSS included 54 items after establishing content validity.

2.2. Construct Validity

The electronic version (i.e., Qualtrics) of the ARSSS was distributed online using Amazon Mechanical Turk (MTurk). The 54-item ARSSS contained 31 visual analog-scale items placed on a 100 mm horizontal line with verbal anchors on the extremes, ranging from strongly disagree to strongly agree. Respondents rated their perceptions by moving the slider to correspond with their level of disagreement (0) or agreement (100). The distance between the marked point (i.e., slider) and the origin of the line is measured to quantify the magnitude of the response. A visual analog scale was chosen because it can provide data that may be treated as interval rather than ordinal level and has been used to assess users' perceptions of AVs [18]. MTurk provided access to a virtual community of workers from different regions of the U.S. with varying backgrounds, who were willing to complete a human intelligence task (HIT). MTurk workers were required to be adults (>18 years old) living in the U.S. and having attempted at least 1000 HITs with a successful completion of at least 95% of their attempted HITs (i.e., Master Workers). A HIT was submitted for USD 5.00, and interested MTurk workers responded using the survey link, which directed them to the Qualtrics ARSSS. Participant responses from 553 adults living in the U.S. were

used to assess the reliability and construct validity (including the factor structure) as part of determining the final psychometric properties of the ARSSS. A conservative sampling approach, based on having 5–10 responses per item ($10 \text{ responses} \times 54 \text{ items} = 540$) and having more than 300 cases, was used to power our analyses [25].

The measurement model was built using a two-stage approach consisting of an EFA followed by a CFA. An EFA was employed to extract the fundamental dimensions of the ARSSS. During EFA, parallel analysis was used to compute eigenvalues from the correlation matrix to determine the number of components to retain for oblimin rotation. The primary goal of factor rotation is to rotate factors within a multidimensional space to arrive at a solution with the best simple structure (i.e., parsimony). This iterative process was repeated until a simple structure was achieved where loadings were maximized on putative factors and minimized on the others. The factor structure from the EFA was then confirmed using a CFA. Hu and Bentler [26] recommend using a relative fit index (i.e., comparative fit index) in combination with an absolute fit index (i.e., root-mean-square error of approximation) as indicators for good fit but caution against over-relying on cutoff indices because it might lead to incorrect rejection of acceptable models.

One hundred participants were asked to complete the ARSSS again after two weeks. To prevent nesting (i.e., due to similar response patterns from the same participant at different time points), the follow-up responses for this group of 100 participants were not entered into the factor analysis.

2.3. Analysis

Data processing was carried out in Rstudio with R version 4.0.4, using the psych and lavaan packages in the tidyverse ecosystem. The measurement model was built using an exploratory factor analysis (EFA) among the 31 visual analog scale items. The other items had different response options and thus could not be analyzed using factor analysis techniques. Item responses that were not selected by any of the 553 respondents were removed from the survey to enhance concision and mitigate respondent burden. An EFA was employed to extract the fundamental dimensions of users' perceptions of transportation options. The EFA was built via the R package lavaan, using the principal axis factoring method and oblimin rotation. The criterion for loading and cross-loading was set at 0.4, and based on this, items were removed from the subscales. Internal consistency and construct reliability were assessed using Cronbach's alpha (α) and composite reliability (McDonald's Ω), respectively, both at a factor level and scale level. Pearson's r and intraclass correlation coefficients (ICC 2,1) were computed to assess the test-retest reliability at the subscale level. A perfect Pearson's correlation of -1 or $+1$ occurs when the variables are perfectly correlated with one another. ICC reliability values can range from 0 to 1 and can be interpreted as poor (<0.75), moderate ($0.75\text{--}0.90$), or good (>0.90 ; [27]). The results (i.e., factor structure) from the EFA informed the factor structure for the CFA. The model was assessed using a range of model fit indices, including root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), comparative fit index (CFI), and the ratio of the chi-square statistic to the respective degrees of freedom (χ^2/df). The cutoff criteria of the model fit indices are detailed below in the analysis.

3. Results

One hundred and ten items were extracted from the literature, and 39 items were generated that were not present in the literature (e.g., ridesharing, ridehailing, autonomous taxis, autonomous shuttles). Items extracted from the technology acceptance literature were adapted to focus on shared AVs. The item pool consisted of 12 double-barreled items, which were split to provide greater clarity. An item pool consisting of 161 items was created and reviewed by all team members. This number of survey items was too large to be included in a functional survey instrument. The research team met and reduced 161 survey items to 54 items. The item pool was shared with sponsors and stakeholders (i.e., Safe Mobility for Life Coalition in Florida, U.S.). Their feedback resulted in the modification of

30 survey items, removal of 4 survey items, addition of 1 survey item, and the inclusion of images of transportation options relevant to the survey.

3.1. Face and Content Validity

During focus groups, the moderator guided the conversation, promoted discussion, and asked follow-up questions. Participant comments were recorded by a member of the research team during the focus groups, which summarized participants' feedback to clarify wording, remove some items, make items clear or concise, and increase the understandability of the survey. During this process, feedback led to the modification of 22 (43%) of the 51 items. Furthermore, definitions of transportation options were modified to align with the survey and the pictures provided in the introductory section of each portion of the survey. Specifically, clarity, concision, complexity, and redundancy were addressed in the revised version of the survey.

The survey faced the following challenges to calculating Flesch Reading Ease Score and Flesch–Kincaid Grade Level Score: (a) This is not a “standard document”—it is a survey, formatted with repeated introductions, required standardized definitions, and required response formats. (b) The topic of the survey itself (“autonomous” and “transportation”) has multiple syllables per word that must be repeated throughout (e.g., the word “autonomous” appears 93 times), along with terminology such as “paratransit.” While all of these multisyllable, higher reading-level words are defined and explained with simpler terminology, the terms themselves remain and are counted towards the overall calculation. After the removal of repeated introductory text and the word “autonomous”, the Flesch–Kincaid Grade Level Score was 8.8, just above the target score of 8, and a reading ease score of 55.7, just below the target score of 60. The Flesch Reading Ease Score and Flesch–Kincaid Grade Level Score are displayed in Table 1.

Table 1. Readability scores of the ARSSS.

Steps	Flesch Reading Ease	Flesch–Kincaid Grade Level
ARSSS ¹	43.0	10.7
ARSSS without repeated introductory text	50.4	9.6
ARSSS without repeated introductory text and without the word “autonomous”	55.7	8.8

¹ ARSSS—Autonomous RideShare Services Survey.

In the first round of review, SMEs rated 50 of the 52 items above the 80% CVI threshold. Item-level CVI scores were 100% (23 items), 90% (14 items), 80% (3 items), and 70% (2 items). Two items were generated in response to SME feedback during the first round. This was done to limit double-barreled items and enhance item clarity. Two newly generated items and two modified items with insufficient item-level CVI scores (i.e., 70%) were sent back to the SMEs for a second round of review. After the second round, the four items surpassed the CVI threshold. In summary, all 54 items were rated above the CVI threshold, resulting in a scale CVI (i.e., mean of the mean item-level CVI scores) of 95%. Feedback from the SME was integrated to refine, reword, and redefine items. This resulted in the refinement (i.e., adding or removing responses, concision) of 15 items and enhanced descriptions of ridesourcing services.

3.2. Construct Validity

The MTurk sample of 553 participants ranged in age from 19 to 71 years old (35.9 + 10.3 years old). The majority of participants were male (66%) and White (71%). This sample included Asian (19%), Black (7%), and other (3%) representation. As a manipulation check, participants rated their familiarity with AVs and transportation services mentioned throughout the survey (Table 2). The most frequent ratings for AVs and autonomous shuttles were somewhat familiar and slightly familiar, respectively.

Table 2. Familiarity with autonomous vehicles and transportation services.

Transportation Type	Not Familiar	Slightly Familiar	Somewhat Familiar	Moderately Familiar	Extremely Familiar
Autonomous vehicles	27 (5%)	141 (26%)	166 (30%)	145 (26%)	74 (13%)
Autonomous taxis	113 (20%)	140 (25%)	128 (23%)	120 (22%)	52 (9%)
Autonomous shuttles	139 (25%)	131 (24%)	119 (22%)	118 (21%)	46 (8%)
Ridesourcing services	21 (4%)	57 (10%)	123 (22%)	190 (34%)	162 (29%)
Ridesharing services	19 (3%)	71 (13%)	147 (27%)	185 (33%)	131 (24%)

A normality check was performed for each item by computing the univariate skewness (<3) and kurtosis (<10 ; [28]). The skew indexes ranged from -0.94 to -0.13 ; the kurtosis indexes ranged from -0.88 to 1.17 . The Kaiser–Meyer–Olkin ($KMO > 0.8$) measure of sampling adequacy suggested that the data seemed appropriate for factor analysis: $KMO = 0.96$. Bartlett’s test of sphericity suggested that there was sufficient significant correlation in the data for an EFA: $\chi^2 (495) = 12,619.65$, $p < 0.001$. The Velicer’s Minimum Average Partial criterion informed the decision to conduct an exploratory factor analysis with four factors.

The results from the Initial EFA (Table 3) displayed signs of low-loading items (<0.4), resulting in item 33 being removed from the survey. The four-factor structure with 30 items, explaining 58.65% of the variance, conceptually represented intention to use, trust, and safety (13 items), potential benefits (7 items), accessibility (7 items), and situation-dependent perceptions (3 items). The factor labels were determined by assessing item content, commonalities, and Loevinger’s coefficient of homogeneity [29].

Table 3. Item-loading from EFA.

Item # Prior to Scale Validation	Factor 1	Factor 2	Factor 3	Factor 4	ARSSS Item #
51	0.83	-	-	-	21
39	0.68	-	-	-	22
48	0.68	-	-	-	30
44	0.66	-	-	-	27
45	0.65	-	-	-	28
42	0.59	-	-	-	25
36	0.59	-	-	-	32
30	0.58	-	-	-	29
34	0.48	-	-	-	26
35	0.43	-	-	-	31
41	0.42	-	-	-	24
40	0.41	-	-	-	23
31	0.41	-	-	-	20
23	-	0.66	-	-	36
49	-	0.65	-	-	39

Table 3. Cont.

Item # Prior to Scale Validation	Factor 1	Factor 2	Factor 3	Factor 4	ARSSS Item #
22	-	0.64	-	-	35
50	-	0.62	-	-	40
24	-	0.56	-	-	38
20	-	0.53	-	-	41
21	-	0.49	-	-	37
27	-	-	0.81	-	44
26	-	-	0.65	-	45
29	-	-	0.58	-	43
25	-	-	0.58	-	42
53	-	-	0.51	-	47
28	-	-	0.48	-	48
52	-	-	0.45	-	46
43	-	-	-	0.72	-
46	-	-	-	0.49	-
32	-	-	-	0.43	-
33	<0.4	<0.4	<0.4	<0.4	-
Internal Consistency	$\alpha = 0.94$	$\alpha = 0.90$	$\alpha = 0.89$	$\alpha = 0.56$	-
Composite Reliability	$\Omega = 0.95$	$\Omega = .89$	$\Omega = 0.89$	$\Omega = 0.56$	-
Average Variance Extracted ¹	AVE = 0.59	AVE = 0.55	AVE = 0.55	AVE = 0.30	-
Variance Explained by each Factor:	23.54%	14.11%	16.05%	4.95%	-

¹ AVE—Average Variance Extracted Maximum values were displayed for item-loading with a cutoff of 0.4.

After the EFA, survey responses of 30 items were assigned to their factor for a confirmatory factor analysis (CFA; Model 1 in Table 4). A multidimensional scale should have five or more items for each factor or subscale [21]. The situational-dependent factor (Factor 4 in Table 3), consisting of 3 items (#32, 43, 46), was not significantly related to any of the other three factors, only explained 4.95% of the overall variance, and was removed from the survey. A second CFA was deployed among 27 items, representing three factors (Model 2 in Table 4). All fit indices exceeded acceptable criteria (Table 4) and improved after the removal of the three items that load on the situational-dependent factor.

Table 4. Model parameters and fit indices from the confirmatory factor analysis.

Model Parameters (N = 553)				Fit Indices		
Models	Factors	Items	χ^2/df	CFI ¹	RMSEA ²	SRMR ³
Model 1	4	30	3.54	0.853	0.073	0.070
Model 2	3	27	3.45	0.881	0.072	0.056
Criteria: Acceptable			<5	>0.8	<0.1	<0.05
Criteria: Good			<3	>0.9	<0.08	<0.1

¹ Comparative Fit Index; ² Root Mean Square Error of Approximation; ³ Standardized Root Mean Square Residual.

Internal consistency of the ARSSS Cronbach's alpha (cutoff: >0.8) and composite reliability (cutoff: >0.7) [30] were used to assess the internal consistency of the items and each of its factors (i.e., after removing factor 4 and items # 32, 43, and 46). Overall, the internal consistency of this scale was excellent (Cronbach's $\alpha = 0.96$), with factors ranging from moderate to excellent (range $\alpha = 0.89$ to 0.94 ; Table 3). The overall Cronbach's α would not be affected by removing any individual items from the scale, as new α 's maintained an α of 0.95 with the deletion of any individual item. Similarly, as shown in Table 3, the composite reliability measures (i.e., construct reliability) for factors 1, 2, and 3 ranged from 0.89 to 0.95 .

A sample of 100 MTurk workers estimated the test-retest reliability of the ARSSS. Participants completed the ARSSS again, two weeks after their first completion of the ARSSS. One extreme outlier (i.e., $>Q3 + 3 \times IQR$ or $<Q1 - 3 \times IQR$) was detected and removed from the analysis. The Bland–Altman plot method was used to visually inspect the test-retest reliability after two weeks (Figure 1). As displayed in Figure 1, 7 of the 99 within-subject test-retest difference scores were outside of the 95% confidence interval $[-16.89, 16.19]$. The total ARSSS scores for test and retest reliability in these 99 participants were significantly and strongly correlated with good reliability ($r = 0.86$, $p < 0.001$, $ICC = 0.99$). The factor scores for test-retest were also significantly and strongly correlated with good reliability: intention to use, trust, and safety ($r = 0.85$, $p < 0.001$, $ICC = 0.99$), potential benefits ($r = 0.70$, $p < 0.001$, $ICC = 0.97$), and accessibility ($r = 0.78$, $p < 0.001$, $ICC = 0.96$). All individual items for the test and retest reliability correlated significantly, with paired sample correlations ranging from 0.59 to 0.70 .

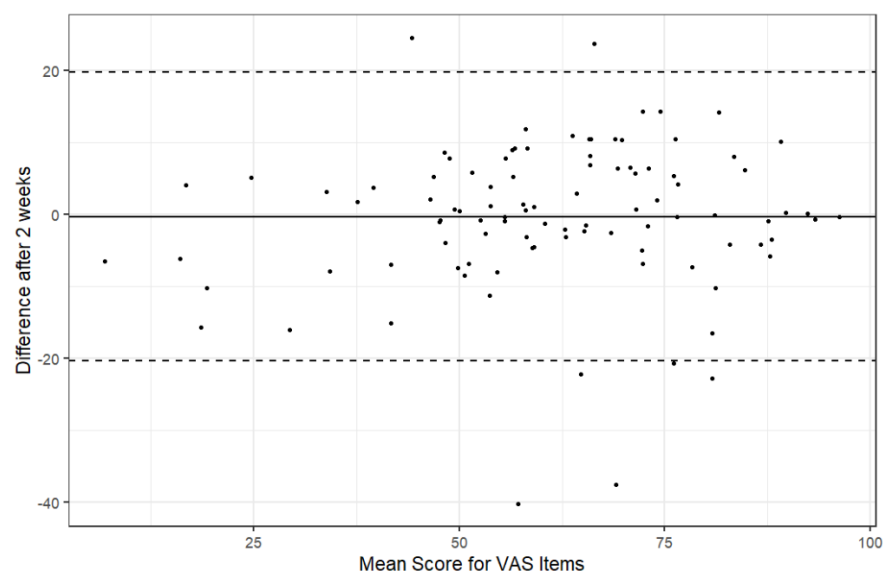


Figure 1. Bland–Altman Plot to indicate the difference between total survey scores after two weeks with the removal of one outlier.

A paper (Supplementary Material) and web-based version (of our survey were constructed by reorganizing items thematically to enhance internal consistency reliability [31].

4. Discussion

The ARSSS was developed to gather demographics, transportation habits, familiarity with AVs and transportation services, and the perceptions of AVs and transportation—especially since the literature showed a gap in offering a similar instrument with good psychometrics. The extant literature informed the development of the initial survey item pool, which was then reduced, and face validity was assessed, followed by establishing content validity via focus groups. The survey psychometrics were established to ensure that the ARSSS is a valid and reliable tool for assessing adults' perceptions of HAVs. The survey

validation and understanding of adults' perceptions of AVs are necessary for informing the acceptance practices among end-users.

Results from this survey may be used to elucidate the effects of AV demonstration projects on users' intention to use, trust, safety, potential benefits, and accessibility of AVs. Demonstration projects are developed and designed with community stakeholders to promote community initiatives and often vary based on region (i.e., climate, rurality), mode of AV (low-speed autonomous shuttle, autonomous van, autonomous taxi, etc.), route (road types, ambient traffic, and routes), and numerous other characteristics and factors [6–8]. While it may be beneficial to compare the effects of different demonstration projects across the U.S., it may also be useful to look at the extent to which perceptions change in one location, before and after adults' exposure to AVs. The ARSSS may be a useful tool to better understand the effects of exposure to AVs for users or for other road users (i.e., drivers, pedestrians) after several months of interacting with the AV during a demonstration project. Other instruments, such as the Automated Vehicle User Perception Survey (AVUPS), may be used with the ARSSS to provide additional insights into adults' perceptions of highly autonomous vehicles. The AVUPS contains three Mokken subscales, which may be used separately or in tandem with one another depending on the research questions or aims. The AVUPS uses the same visual analog scale as the ARSSS and, if administered together, may reduce respondent confusion or exhaustion due to not requiring respondents to switch between different scales. However, the AVUPS does not collect demographics, transportation habits, or perceptions pertaining to autonomous ridehailing or ridesharing services. Thus, results from the ARSSS may provide a more holistic understanding of the road user, their transportation preferences, and available transit options, as well as their perceptions of novel and emerging modes of transportation, such as autonomous ridehailing and ridesharing services.

Results from demonstration projects cannot be directly compared as they used different conceptual frameworks, surveys, routes, automated road transport systems, and private vs. shared services. For example, demonstration projects have used a Tesla Model X [32], low-speed autonomous shuttle in a 10 min closed loop without ambient traffic in the U.S. [6], a retrofitted Ford Transit operating on the interstate, mixed traffic, and gravel roads in the rural U.S. on a 50-mile loop [33], and six automated shuttles operating on a dedicated lane on a 2.5 km route in Greece [34]. Although there are numerous variables to consider, the use of a validated survey may be the first step in making comparisons between demonstration projects. However, for now, survey validation supports results within a demonstration project to assess pre- and post-exposure differences in users' perceptions of AVs in the U.S. The ARSSS was validated using a U.S. sample that varied by age and sex but not by country of origin or residence. Thus, a limitation of the ARSSS is that the survey results may not be reliable nor valid if administered among an international sample. This limitation provides an excellent opportunity for collaboration across government agencies, universities, and other international institutions that are interested in the public's opinion of autonomous ridesharing services and AVs.

This survey lays an important foundation in assessing perceptions and acceptance of autonomous ridesharing services—which is not a guarantee for the actual acceptance of ridesharing. In other words, intent does not always lead to behavior. Further empirical investigations are necessary to provide substantive evidence for ensuring actual acceptance practices versus perceptions thereof. Future projects may consider measuring individuals with diverse demographics, e.g., people with disabilities, mobility vulnerable, socioeconomic disadvantaged, or those living in rural areas. Moreover, other factors such as technology literacy, culture, evolution of the technology, private vs. public ridesharing, and complex environments (e.g., presence of gravel or ice) that are context- and situation-dependent must be considered in the item pool.

The survey development detailed in this paper encompassed a multi-pronged approach to examine, quantify, and refine the psychometrics of the ARSSS. The survey was enhanced through an iterative process that eventually led to collecting responses

using MTurk. MTurk afforded researchers a quick, efficient, and reliable method to collect 553 responses within 48 h. Several acceptance models, focus groups, stakeholders, subject-matter experts, and measurement approaches were used to inform the development of the ARSSS. To our knowledge, the ARSSS is the first validated tool to measure user perceptions pertaining to autonomous ridesharing and ridesourcing services.

The next steps include using this survey in demonstration projects across multiple sites in Florida. An automated shuttle will be deployed across these regions, and the ARSSS will be used to quantify perceptions before and after the shuttle ride. This will allow our research team to compare different regions and their intention to use automated shuttles in an automated road transport system. Future survey validation will include assessing the convergent validity with surveys being used across the globe, as well as criterion validity based on end-user acceptance and actual use of the technology. Last but not least, additional survey validation may provide a brief version of this survey, improving the feasibility of survey administration and enhancing the fit indices.

5. Conclusions

The approach adopted in this study ensured that the survey instrument design included items with adequate face validity and content validity. This validation supports the claim that the ARSSS can be used to quantify the relevant theoretical constructs pertaining to users' perceptions and acceptance of AVs. Furthermore, face validation suggests that the intended audience believes the items are clear, appropriate, and relevant to transportation and AVs [19]. The construct validity and test-retest validity used an efficient approach, via MTurk, to determine and establish the psychometrics. Specifically, the EFA and CFA confirmed a three-factor structure meaning that the visual analog scaled items load on one of three factors. Survey results from the three factors, Intention to use, trust, and safety (13 items), Potential benefits (7 items), and Accessibility (7 items), can be used to inform policy makers, researchers, and manufacturers so that they are better able to understand adults' perceptions of AVs and how these factors change after experiencing the technology. The ARSSS may be used free of charge and is made available for administration via pencil/paper, tablet, or online. Survey responses can be aggregated from the visual analog scale and can be displayed as percentages, given that the scale ranges from 0 to 100. The ARSSS lays the foundation for a valid and reliable survey that may be adapted for other groups, including persons with disabilities, the mobility vulnerable, people in rural areas, and the transportation disadvantaged.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/futuretransp3020042/s1>, File S1: The Autonomous RideShare Services Survey.

Author Contributions: Conceptualization, S.C. and J.M.; methodology, S.C. and J.M.; formal analysis, J.M.; investigation, S.C. and J.M.; resources, S.C. and J.M.; data curation, S.C. and J.M.; writing—original draft preparation, S.C. and J.M.; writing—review and editing, S.C. and J.M.; visualization, J.M.; supervision, S.C. and J.M.; project administration, S.C. and J.M.; funding acquisition, S.C. and J.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Florida Department of Transportation, grant number BDV31 977-128, and the APC was waived.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of the University of Florida (IRB201903309-approved 3 September 2019).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data will be made available upon request without reservation.

Acknowledgments: The authors acknowledge and thank the Florida Department of Transportation (FDOT) and the Institute for Driving, Activity, Participation, and Technology (I-DAPT) for providing financial support and materials for this project. Special thanks to Michael Scicchitano and Tracy Johns at the Florida Survey Research Center for their expertise and support and to FDOT Project Manager Gail Holley and the Safe Mobility for Life Coalition for their contributions in terms of their expertise, experience, and constructive feedback throughout the scope of this project.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Society of Automotive Engineers. *Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems*; SAE International: Warrendale, PA, USA, 2016.
2. National Highway Traffic Safety Administration. *Automated Driving Systems 2.0: A Vision for Safety*; U.S. Department of Transportation: Washington, DC, USA, 2017.
3. Narayanan, S.; Chaniotakis, E.; Antoniou, C. Shared autonomous vehicle services: A comprehensive review. *Transp. Res. Part C Emerg. Technol.* **2020**, *111*, 255–293. [\[CrossRef\]](#)
4. Gurumurthy, K.; Kockleman, K. Analyzing the dynamic ride-sharing potential for shared autonomous vehicle fleets using cellphone data from Orlando, Florida. *Comput. Environ. Urban Syst.* **2018**, *71*, 177–185. [\[CrossRef\]](#)
5. Stocker, A.; Shaheen, S. *Shared Automated Mobility: Early Exploration and Potential Impacts*, Road Vehicle Automation 4; Lecture Notes in Mobility; Springer: Cham, Switzerland, 2017.
6. Classen, S.; Mason, J.; Hwangbo, S.-W.; Wersal, J.; Rogers, J.; Sisiopiku, V. Older drivers' experience with automated vehicle technology. *J. Transp. Health* **2021**, *22*, 101107. [\[CrossRef\]](#)
7. Nordhoff, S.; Winter, J.; Payre, W.; Arem, B.; Happee, R. What impressions do users have after a ride in an automated shuttle? An interview study. *Transp. Res. Part F Traffic Psychol. Behav.* **2019**, *63*, 252–269. [\[CrossRef\]](#)
8. Hoff, K.A.; Bashir, M. Trust in automation: Integrating empirical evidence on factors that influence trust. *Hum. Factors* **2015**, *57*, 407–434. [\[CrossRef\]](#)
9. Davis, F. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Manag. Inf. Syst. Q.* **1989**, *13*, 319–340. [\[CrossRef\]](#)
10. Osswald, S.; Wurhofer, D.; Trösterer, S.; Beck, E.; Tscheligi, M. Predicting information technology usage in the car: Towards a car technology acceptance model. In Proceedings of the AutomotiveUI '12, 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Portsmouth, NH, USA, 17–19 October 2012.
11. Venkatesh, V.; Morris, M.G.; Hall, M.; Davis, G.B.; Davis, F.D.; Walton, S.M. User Acceptance of Information Technology: Toward a Unified View. *Manag. Inf. Syst. Q.* **2003**, *27*, 425–478. [\[CrossRef\]](#)
12. Mason, J.; Classen, S.; Sisiopiku, J.W.A.V. Construct validity and test–retest reliability of the automated vehicle user perception survey. *Front. Psychol.* **2021**, *12*, 626791. [\[CrossRef\]](#)
13. Nordhoff, S.; van Arem, B.; Happee, R. Conceptual model to explain, predict, and improve user acceptance of driverless podlike vehicles. *Transp. Res. Rec.* **2016**, *2602*, 60–67. [\[CrossRef\]](#)
14. Thomas, E.; McCrudden, C.; Wharton, Z.; Behera, A. The perception of autonomous vehicles by the modern society: A survey. *IET Intell. Transp. Syst.* **2020**, *14*, 1228–1239. [\[CrossRef\]](#)
15. Casley, L.; Jardim, A.; Quartulli, A. *Study of Public Acceptance of Autonomous Cars Interactive Qualifying Project*; Worcester Polytechnic Institute: Worcester, UK, 2014.
16. Das, S.; Dutta, A.; Fitzpatrick, K. Technological perception on autonomous vehicles: Perspectives of the non-motorists. *Technol. Anal. Strateg. Manag.* **2020**, *32*, 1335–1352. [\[CrossRef\]](#)
17. Othman, K. Public acceptance and perception of autonomous vehicles: A comprehensive review. *AI Ethics* **2021**, *1*, 355–387. [\[CrossRef\]](#)
18. Mason, J.; Classen, S.; Wersal, J.; Sisiopiku, V. Establishing face and content validity of a survey to assess users' perceptions of automated vehicles. *Transp. Res. Rec.* **2020**, *2674*, 538–571. [\[CrossRef\]](#)
19. Holden, R.B. Face Validity. In *The Corsini Encyclopedia of Psychology*, 4th ed.; Weiner, I.B., Craighead, W.E., Eds.; Wiley: Hoboken, NJ, USA, 2010. [\[CrossRef\]](#)
20. Colton, D.; Covert, R.W. *Designing and Constructing Instruments for Social Research and Evaluation*; John Wiley & Sons: Hoboken, NJ, USA, 2007.
21. DiStefano, C.; Zhu, M.; Mindril, D. Understanding and using factor scores: Considerations for the applied researcher. *Pract. Assess. Res. Eval.* **2009**, *14*, 20.
22. Knekta, E.; Runyon, C.; Eddy, S. One size doesn't fit all: Using factor analysis to gather validity evidence when using surveys in your research. *CBE—Life Sci. Educ.* **2019**, *18*, rm1. [\[CrossRef\]](#)
23. Owlsey, C.; Stalvey, B.; Wells, J.; Sloane, M. Older drivers and cataract: Driving habits and crash risk. *J. Gerontol. Ser. A* **1999**, *54*, 203–211.

24. Lynn, M. Determination and Quantification of Content Validity. *Nurs. Res.* **1986**, *35*, 382–385. [[CrossRef](#)]
25. Bryant, F.B.; Yarnold, P.R. *Principal-Components Analysis and Exploratory and Confirmatory Factor Analysis*; American Psychological Association: Washington, DC, USA, 1995.
26. Hu, L.; Bentler, P. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equ. Model. A Multidiscip. J.* **1999**, *6*, 1–55. [[CrossRef](#)]
27. Fleiss, J.; Levin, B.; Paik, M. Statistical methods for rates. In *Statistical Methods for Rates and Proportions*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2013.
28. Kline, R.B. Principles and Practice of Structural Equation Modeling. In *Methodology in Social Sciences*; The Guilford Press: New York, NY, USA, 2010.
29. Fabrigar, L.R.; MacCallum, R.C.; Wegener, D.T.; Strahan, E.J. Evaluating the use of exploratory factor analysis in psychological research. *Psychol. Methods* **1999**, *4*, 272–299. [[CrossRef](#)]
30. Hair, J.F.; Black, W.C.; Babin, B.J.; Anderson, R.E.; Tatham, R.L. *Multivariate Data Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 1998.
31. Melnick, S.A. The effects of item grouping on the reliability and scale. *Educ. Psychol. Meas.* **1993**, *53*, 211–216. [[CrossRef](#)]
32. Bousonville, T.; Rosler, I.; Vogt, J.; Wolniak, N. Performance and acceptance of a partially automated shuttle service for commuters using a Tesla Model X. *Transp. Res. Procedia* **2022**, *64*, 98–106. [[CrossRef](#)]
33. Mason, J.; Carney, C.; Gaspar, J. Autonomous shuttle operating on highways and gravel roads in rural America: A demonstration study. *Geriatrics* **2022**, *7*, 140. [[CrossRef](#)] [[PubMed](#)]
34. Nordhoff, S.; Madigan, R.; Arem, B.V.; Merat, N.; Happee, R. Interrelationships among predictors of automated vehicle acceptance: A structural equation modelling approach. *Theor. Issues Ergon. Sci.* **2021**, *22*, 383–408. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.