

Table C-9. Psychometric data reported in studies included for Key Question 1C

Study	Measure and Psychometric Property	Description of Analysis	Result
Finger et al. 2014 ⁶	Modified Impact of Vision Impairment (IVI) questionnaire, Construct Validity	Response category thresholds (could participants discriminate between items) (between 0.7 and 1.3 is acceptable).	Response category thresholds: The original version had 3 items misfit. After determining the 2 domains, these 3 items were removed, along with 6 other items. The resulting 28-item instrument had no misfit items.
Finger et al. 2014 ⁶	Modified Impact of Vision Impairment (IVI) questionnaire, Face validity	Focus group discussions and telephone interviews with vision-impaired patients, healthy controls, and professionals.	Focus group discussions reduced an initial item pool from 76 to 52 items. Then 198 legally blind people were interviewed by telephone to reduce the 52 to 37. Eliminations were based on a floor effect (tasks too easy to bother asking about). Rephrasing of questions, and changing of response options.
Finger et al. 2014 ⁶	Modified Impact of Vision Impairment (IVI) questionnaire, Reliability	Separation reliability, PSI, which is how well the instrument classified respondents into different levels of the trait (PSI >2 is considered acceptable).	PSI: Met their criteria for acceptability.
Finger et al. 2014 ⁶	Modified Impact of Vision Impairment (IVI) questionnaire, Reliability	Internal consistency reliability. Person reliability, which reflects the spread of the underlying trait in the sample (person reliability >0.8 is considered acceptable).	Person reliability: Met their criteria for acceptability.
Finger et al. 2014 ⁶	Modified Impact of Vision Impairment (IVI) questionnaire, Reliability	Internal consistency reliability, same as above.	IR: Met their criteria for acceptability
Finger et al. 2014 ⁶	Modified Impact of Vision Impairment (IVI) questionnaire, Construct Validity	A test of unidimensionality based on the residuals of the first factor in principal components analyses (>50% is considered acceptable).	Unidimensionality: Residuals of first factor: Met their criteria for acceptability. There was no evidence of multidimensionality.
Finger et al. 2014 ⁶	Modified Impact of Vision Impairment (IVI) questionnaire, Construct Validity	Another test of unidimensionality, based on the first contrast of residuals (<2.5 eigenvalues is considered acceptable).	Unidimensionality: eigenvalues of first contrast: Met their criteria for acceptability. There was no evidence of multidimensionality.
Finger et al. 2014 ⁶	Modified Impact of Vision Impairment (IVI) questionnaire, Construct Validity	Targeting, which is whether the items adequately target the ability of respondents (a difference of less than 1.0 logits is considered acceptable).	Targeting: "slightly suboptimal", but "still well within acceptable levels."
Finger et al. 2014 ⁶	Modified Impact of Vision Impairment (IVI) questionnaire, Construct Validity	Differential item functioning, which is whether sample subgroups with similar underlying ability (e.g., of different age, sex, etc.) have different scores on the instrument.	Differential Item Functioning: None of the items showed this. Neither ADLMS nor EWB subscores were statistically significantly correlated with age, sex, marital status, living situation, employment, or education (see Table 3 of the article).

Table C-9. Psychometric data reported in studies included for Key Question 1C (continued)

Study	Measure and Psychometric Property	Description of Analysis	Result
Finger et al. 2014 ⁶	Modified Impact of Vision Impairment (IVI) questionnaire, Construct Validity	Whether their responses are associated with their eye conditions	Association with eye conditions (see Table 3 of the article). The ADLMS subscores varied among patients with different diagnoses, but the EWB subscores did not. For ADLMS, the means for RP, AMD, OR, glaucoma, and other were 0.15, -0.27, -0.37, -0.19, and 0.19, respectively (overall $p=0.018$). Thus the poorest ADLMS was found among those with AMD and glaucoma. For EWB, the means for RP, AMD, OR, glaucoma, and other were 0.24, 0.21, 0.45, 0.11, and 0.48 (overall $p=0.685$).
Finger et al. 2014 ⁶	Modified Impact of Vision Impairment (IVI) questionnaire, Construct Validity	Whether their responses are associated with other health aspects.	Association with other health problems. Both ADLMS and EWB subscores correlated with 4 other measures of health (general health, other health problems, do other health problems interfere, and anxiety/depression). As expected, higher ADLMS and EWB subscores were predictive of better health responses to these questions.
Finger et al. 2014 ⁷	Very Low Vision Instrumental Activities of Daily Living (IADL-VLV), Face validity	Focus group discussions and telephone interviews with patients with impaired vision, healthy controls, and professionals.	Authors began with 296 items from existing ADL tools. These were decreased to 25 general activities based on importance rankings with 62 participants with severe low vision. A panel of low vision experts then reduced the 25 activities to 11, which were comprised of 53 specific tasks. Tasks included "table and shelf searches for cutlery and crockery items, clock face and symbol recognition, signature placement, drink pouring, clothes sorting, and the understanding of hand gestures" (see Table 1 of the article for a complete list of the 53 tasks). The 53 tasks were all submitted to construct validity testing among 40 study participants.
Finger et al. 2014 ⁷	Very Low Vision Instrumental Activities of Daily Living (IADL-VLV), Construct Validity	Response category thresholds (could participants discriminate between items) (between 0.7 and 1.3 is acceptable).	Response category thresholds: The original version (53 tasks) had 5 misfit items. A second version (27 tasks) had 2 misfit items. The final version (23 tasks) had 0 misfit items.
Finger et al. 2014 ⁷	Very Low Vision Instrumental Activities of Daily Living (IADL-VLV), Reliability	Separation reliability, PSI, which is how well the instrument classified respondents into different levels of the trait (PSI >2 is considered acceptable).	PSI: Final version met their criteria for acceptability. Both the 1st and 2nd versions also met their criteria.
Finger et al. 2014 ⁷	Very Low Vision Instrumental Activities of Daily Living (IADL-VLV), Reliability	Internal consistency reliability. Person reliability, which reflects the spread of the underlying trait in the sample (person reliability >0.8 is considered acceptable).	Person reliability: Final version met their criteria for acceptability. Both the 1st and 2nd versions also met their criteria.

Table C-9. Psychometric data reported in studies included for Key Question 1C (continued)

Study	Measure and Psychometric Property	Description of Analysis	Result
Finger et al. 2014 ⁷	Very Low Vision Instrumental Activities of Daily Living (IADL-VLV), Construct Validity	A test of unidimensionality based on the residuals of the 1st factor in principal components analyses (>50% is considered acceptable).	Unidimensionality: Variance of first factor: Initial version (53 tasks) was 70.7%. Second version (27 tasks) was 71.6%, and the final version (23 tasks) was 77.5%.
Finger et al. 2014 ⁷	Very Low Vision Instrumental Activities of Daily Living (IADL-VLV), Construct Validity	Another test of unidimensionality, based on the 1st contrast of residuals (<2.5 eigenvalues is considered acceptable).	Unidimensionality: eigenvalues of 1st contrast: The initial version had severe multidimensionality (eigenvalue for the first contrast 7.1). Even the final version did not meet their criteria for acceptability (eigenvalue for the 1st contrast 4.7). This was because there were 5 underlying factors, not just 1. Authors did not attempt to list the 5, but they did state that the 23 final tasks were related to 6 activities: table search, recognition of symbols, clock reading, signature placement, clothes sorting, and recognition of hand gestures.
Finger et al. 2014 ⁷	Very Low Vision Instrumental Activities of Daily Living (IADL-VLV), Construct Validity	Whether their responses are associated with age or sex (which they ideally would not be).	Association with age/sex. After controlled for cognitive impairment and depression, there was no association between scores on the final instrument and age or sex. However, they are associated with both cognitive impairment and depression.
Bittner et al. 2011 ²	Grating contrast sensitivity (GCS), Construct Validity	Separately for the 8 patients with RP and the 12 with OR, authors computed the correlation between the newly developed GCS and the standard and "well-validated" test called the Pelli-Robson chart. Perfect validity would be indicated by (1) strong correlation, (2) a slope of 1.0 and (3) an intercept of 0.	GCS did not demonstrate validity for patients with either RP or OR (Figure 4 in the article). Using the Pelli-Robson test as the gold standard, GCS overestimated patients' contrast sensitivity. For some patients with RP, it was not possible to obtain values using the gold standard. Even leaving out those patients, GCS overestimated contrast sensitivity.
Bittner et al. 2011 ²	Grating contrast sensitivity (GCS), Reliability	Test-retest reliability, same as above	For RP, the median test-retest CR _{.95} of GCS was 0.13 for within-visit and 0.15 for between-visit (log-unit scale, see Figure 5 in the article). For OR patients with vision worse than 20/350, the median test-retest CR _{.95} of GCS was 0.13 for within-visit and 0.34 for between-visit (log-unit scale, see Figure 5 in the article). For OR patients with vision better than 20/350, the median test-retest CR _{.95} of GCS was 0.15 for within-visit and 0.41 for between-visit (log-unit scale, see Figure 5 in the article).
Bittner et al. 2011 ²	Pelli-Robson (PR) contrast sensitivity test, Reliability	Test-retest reliability, same as above	For RP, the median test-retest CR _{.95} of PR was 0.14 for between-visit (log-unit scale, see Figure 5 in the article). For OP patients, it was 0.24.
McKnight and Babcock-Parziale 2007 ⁸	FAST (Functional Assessment of Self-Reliance on Tasks), clinician-completed Responsiveness	Responsiveness, as measured by the difference in scores between admission to the program and discharge from the program.	As shown in Figure 2 of the article, discharge scores were consistently much higher than admission scores. Authors did not quantify the size of the difference.

Table C-9. Psychometric data reported in studies included for Key Question 1C (continued)

Study	Measure and Psychometric Property	Description of Analysis	Result
McKnight and Babcock-Parziale 2007 ⁸	FAST (Functional Assessment of Self-Reliance on Tasks), patient-completed Responsiveness	Responsiveness, as measured by the difference in scores between admission to the program and discharge from the program.	As shown in Figure 2 of the article, discharge scores were consistently much higher than admission scores. Authors did not quantify the size of the difference.
McKnight and Babcock-Parziale 2007 ⁸	FAST (Functional Assessment of Self-Reliance on Tasks), clinician-completed, vs. patient-completed, Construct validity	Whether the difficulty of items (as judged by Rasch analysis) was similar for clinician-completed vs. patient-completed forms	As shown in Figure 1 of the article, there was a near linear relationship between the two versions of FAST with respect to item difficulty. The single exception was reading, which was judged to be easier for patients when clinicians judged it as compared to when patients judged it.
McKnight and Babcock-Parziale 2007 ⁸	FAST (Functional Assessment of Self-Reliance on Tasks), clinician-completed, vs. patient-completed, Construct validity	Whether patients' abilities (as judged by Rasch analysis) were similar for clinician-completed vs. patient-completed forms	When the clinician-completed form was used to try to predict the patient-completed form, there was a relatively weak relationship. The slope was only 0.35, and only 55% of the variance in patient scores were explained by clinician scores. However, a multiple regression (Table 1 of the article) found that the timing of administration (at admission or at discharge) was the primary explanatory factor. These data indicate moderate construct validity for the 2 versions of the forms.
Roman et al. 2007 ⁹	Light perception test: full field flash test, Reliability	Test-retest reliability, whether patients' light sensitivities (as measured by FST-2) were similar within a given visit	Good test-retest reliability. The standard deviation of within-visit sensitivity was 1.41 dB, and the median (estimated using Figure 2b of the article) appears to be approximately 17 dB.
Roman et al. 2007 ⁹	Light perception test: full field flash test, Construct validity	Whether patients' abilities were measured to be worse than patients with normal vision	Good construct validity. Those with normal vision had a mean sensitivity of 61.5 dB, whereas patients had a median (estimated using Figure 2b of the article) of approximately 17 dB.
Roman et al. 2007 ⁹	Light perception test: full field flash test, Construct validity	Whether patients' abilities as measured by the latest device iteration (FST2) was similar to the abilities measured by the previous device iteration (FST1)	Good construct validity. The correlation between FST1 and FST results was $r=0.98$, slope=1.0, intercept=0.8dB. This number included both patients and normal, Figure 2c of the article shows that the correlation was strong for patients themselves.
Kiser et al. 2006 ¹⁰	Light perception test: dark adaptometry, Reliability	Test-retest reliability. The metric was the CoV, which is the SD of the time required to reach the person's light perception threshold divided by the average time the person required to reach the threshold. CoV is on the percentage scale, and lower numbers indicate greater test-retest reliability.	Only 16 of 33 RP patients could actually complete this test. All 32 other patients with other eye conditions could complete it. Most patient groups averaged about CoV of 10% to 20% (see Figure 3 in the article for curves for separate groups of patients). Authors did not report means for the different groups or across groups.

Table C-9. Psychometric data reported in studies included for Key Question 1C (continued)

Study	Measure and Psychometric Property	Description of Analysis	Result
Kiser et al. 2006 ¹⁰	Light perception test: dark-adapted Humphrey perimetry: rod-based sensitivity using blue-green stimuli (500 nm), Reliability	Test-retest reliability. Authors computed each patient's CR _{.95} . This was done both between-visit and within-visit. A low CR _{.95} indicates good test-retest reliability. CR _{.95} is on the same scale as the perimetry testing, which is dB.	Only 15 of 33 patients with RP could both do this test and provide sensible results. For patients with other conditions, 19 of 32 could both do the test and provide sensible results. For available results for RP-I patients, the median CR _{.95} was 5 for between-visit and 1.5 for within-visit. For available results for RP-II patients, the CR _{.95} was 1 for between-visit and 1 for within-visit. For available results for RP-III patients, the CR _{.95} was 2 for between-visit and 2 for within-visit. For available results for MD-I patients, the CR _{.95} was 5.5 for between-visit and 3.5 for within-visit. For available results for OR patients, the CR _{.95} was 6 for between-visit and 2.5 for within-visit.
Kiser et al. 2006 ¹⁰	Light perception test: dark-adapted Humphrey perimetry: cone-based sensitivity using red stimuli (650 nm), Reliability	Test-retest reliability, same as above	Only 15 of 33 patients with RP could both do this test and provide sensible results. For patients with other conditions, 19 of 32 could both do the test and provide sensible results. For available results for RP-I patients, the median CR _{.95} was 3 for between-visit and 1 for within-visit. For available results for RP-II patients, the median CR _{.95} was 4.5 for between-visit and 1.5 for within-visit. For available results for RP-III patients, the median CR _{.95} was 6 for between-visit and 2 for within-visit. For available results for MD-I patients, the CR _{.95} was 10 for between-visit and 2 for within-visit. For available results for OR patients, the CR _{.95} was 5.5 for between-visit and 2.5 for within-visit.
Kiser et al. 2006 ¹⁰	Light perception test: dark-adapted Humphrey perimetry: rod-cone sensitivity ratios, Reliability	Test-retest reliability, same as above	Only 15 of 33 patients with RP could both do this test and provide sensible results. For patients with other conditions, 19 of 32 could both do the test and provide sensible results. For available results for RP-I patients, the CR _{.95} was 3 for between-visit and 2 for within-visit. For available results for RP-II patients, the CR _{.95} was 5 for between-visit and 3 for within-visit. For available results for RP-III patients, the CR _{.95} was 5 for between-visit and 5 for within-visit. For available results for MD-I patients, the CR _{.95} was 6.5 for between-visit and 5 for within-visit. For available results for OR patients, the CR _{.95} was 5 for between-visit and 2.5 for within-visit.

Table C-9. Psychometric data reported in studies included for Key Question 1C (continued)

Study	Measure and Psychometric Property	Description of Analysis	Result
Kiser et al. 2006 ¹⁰	Light perception: Full-field flash test, Reliability	Test-retest reliability, same as above	All but 2 of 77 patients could perform this test and provide sensible results (authors did not report the eye conditions of these 2). RP-I patients had a mean of 43 dB with a CR _{.95} of 6 dB. This means that a typical RP-I patient had a threshold of 43 dB, and one would expect with 95% probability that a retest would be between 37dB and 49 dB. RP-II patients had a mean of 39 dB with a CR _{.95} of 7dB. RP-III patients had a mean of 26 dB with a CR _{.95} of 9 dB. RP-IV patients had a mean of 19 dB with a CR _{.95} of 12dB. MD-I patients had a mean of 60 dB with a CR _{.95} of 8dB. MD-II patients had a mean of 64 dB with a CR _{.95} of 6 dB. OR patients had a mean of 50 dB with a CR _{.95} of 10dB. DR patients had a mean of 48 dB with a CR _{.95} of 16 dB.
Kiser et al. 2006 ¹⁰	Light perception test: dark adaptometry and full-field flash test, Construct Validity	Correlation between results of dark adaptometry (threshold dB) and full-field flash test (threshold dB)	The correlation was only r=0.37, and the slope was 2.6, which clearly indicates that the 2 tests are measuring different traits. Authors theorized that the problem was that adaptometry was limited by the device ("limited response range of the SST") and caused a ceiling effect, which "limits the thresholds compared with those of the full-field flash test."
Kiser et al. 2006 ¹⁰	Light perception test: dark-adapted Humphrey perimetry and full-field flash test, Construct Validity	Correlation between results of full-field flash test (threshold dB) and Humphrey perimetry	The correlation was 0.60 and the slope was -1.42. Authors theorized that the problem was MD-I patients were hampering the analysis. When they excluded MD-I patients, the correlation rose to 0.8 and the slope became -1.31.
Babcock-Parziale et al. 2005 ¹¹	VA-13, Responsiveness	Responsiveness. A comparison of Rasch-based person abilities, pretreatment vs. post-treatment	Patients had improved about 0.63 logit, which is less than typically observed in this field (2 or even 4 logits according to the authors), and so the authors stated that VA-13 instrument was under-responsive. Authors noted a ceiling effect in VA-13 responses.
Babcock-Parziale et al. 2005 ¹¹	VA-13, Face validity	Whether the distribution of pre-treatment item difficulty (Rasch-based analysis) was "the same order of difficulty that is observed in clinical practice at admission or in pre-test self-reports."	Two specific items were disordered: according to the VA-13, reading of newspapers/magazine was easier for patients than reading mail, however according to the authors' clinical expertise, the reverse is true. All remaining items were ordered as the authors expected. Thus 11/13 items achieved the expected ordering.
Babcock-Parziale et al. 2005 ¹¹	VA-13, Construct Validity	A comparison of Rasch-based item difficulties abilities, pretreatment vs. post-treatment	Item difficulty would not be expected to change pre vs post. Only 2 of 13 items seem to have changed in difficulty after treatment (reading mail, which became easier, and watching TV, which became more difficult) (see Figure 1 of the article). The authors interpreted this as evidence for the construct validity of the VA-13.
Babcock-Parziale et al. 2005 ¹¹	VA-13, Construct Validity	Response category thresholds: authors considered values between 0.6 and 1.4 as acceptable	Response category thresholds: data for both pre-test and post-test met the authors' criteria for acceptability.

Table C-9. Psychometric data reported in studies included for Key Question 1C (continued)

Study	Measure and Psychometric Property	Description of Analysis	Result
Babcock-Parziale et al. 2005 ¹¹	VA-13, Reliability	Internal consistency reliability as measured by Cronbach's alpha. Item reliability.	For items, Cronbach's alpha was 0.81 for the retrospective pre-test and 0.76 for the post-test, indicating good internal consistency reliability.
Babcock-Parziale et al. 2005 ¹¹	VA-13, Reliability	Internal consistency reliability as measured by Cronbach's alpha. Person reliability.	For person reliability estimates, Cronbach's alpha was 0.71 for the retrospective pre-test and 0.27 for the post-test. This latter value of 0.27 was deemed by the authors to be poor.
Babcock-Parziale et al. 2005 ¹¹	VA-13, Reliability	Separation reliability (2 or more is considered acceptable).	For the retrospective pre-test this was only 1.57, and for the post-test it was only 0.60. The authors deemed these values unacceptably low.
Babcock-Parziale et al. 2005 ¹¹	FAST (Functional Assessment of Self-Reliance on Tasks), Responsiveness	Responsiveness. A comparison of Rasch-based person abilities, pre-treatment vs post-treatment	Acceptably responsive. The logit change was of 2.5 logits corresponded to a large effect size of $d=1.8$.
Babcock-Parziale et al. 2005 ¹¹	FAST (Functional Assessment of Self-Reliance on Tasks), Face validity	Whether the distribution of pre-treatment item difficulty (Rasch-based analysis) was "the same order of difficulty that is observed in clinical practice at admission or in pre-test self-reports."	Distribution of item difficulties was consistent with the authors' opinion.
Babcock-Parziale et al. 2005 ¹¹	FAST (Functional Assessment of Self-Reliance on Tasks), Construct Validity	A comparison of Rasch-based item difficulties abilities, pre-treatment vs. post-treatment	Item difficulty would not be expected to change pre vs post. Only 3 of 11 items seem to have changed in difficulty after treatment (reading, which became easier, and home maintenance and fine motor dexterity, both of which became harder) (see Figure 3 of the article). The authors interpreted this as evidence for the construct validity of the VA-13.
Babcock-Parziale et al. 2005 ¹¹	FAST (Functional Assessment of Self-Reliance on Tasks), Construct Validity	Response category thresholds: authors considered values between 0.6 and 1.4 as acceptable	Response category thresholds: data for both pre-test and post-test met the authors' criteria for acceptability
Babcock-Parziale et al. 2005 ¹¹	FAST (Functional Assessment of Self-Reliance on Tasks), Reliability	Internal consistency reliability as measured by Cronbach's alpha. Item reliability	For items, Cronbach's alpha was 0.97 for the pre-test and 0.95 for the post-test, indicating good internal consistency reliability for items.
Babcock-Parziale et al. 2005 ¹¹	FAST (Functional Assessment of Self-Reliance on Tasks), Reliability	Internal consistency reliability as measured by Cronbach's alpha. Person reliability	For person reliability estimates, Cronbach's alpha was 0.90 for the pre-test and 0.85 for the post-test, indicating good internal consistency reliability for person abilities.
Babcock-Parziale et al. 2005 ¹¹	FAST (Functional Assessment of Self-Reliance on Tasks), Reliability	Separation reliability (2 or more is considered acceptable).	For pre-test the value was 2.9, and for post-test it was 2.37. Both meet criteria for acceptability.

Table C-9. Psychometric data reported in studies included for Key Question 1C (continued)

Study	Measure and Psychometric Property	Description of Analysis	Result
Kiser et al. 2005 ⁴	Pelli-Robson (PR) contrast sensitivity test, regular, Reliability	Test-retest reliability, same as above	Median values of CR _{.95} were: 0.30 for RP-I, 0.31 for RP-II, 0.49 for RP-III, 0.48 for MD-I, 0.47 for MD-II, 0.46 for DR, and 0.19 for OR. See Figure 4 of the article
Kiser et al. 2005 ⁴	Pelli-Robson (PR) contrast sensitivity test, dim, Reliability	Test-retest reliability, same as above	Median values of CR _{.95} were: 0.22 for RP-I, 0.50 for RP-II, 0.38 for RP-III, 0.58 for MD-I, 0.27 for MD-II, 0.30 for DR, and 0.30 for OR. See Figure 4 of the article
Kiser et al. 2005 ⁴	Pelli-Robson (PR) contrast sensitivity test, glare, Reliability	Test-retest reliability, same as above	Median values of CR _{.95} were: 0.25 for RP-I, 0.68 for RP-II, 0.10 for RP-III, 0.59 for MD-I, 0.58 for MD-II, 0.30 for DR, and 0.47 for OR. See Figure 4 of the article
Stelmack et al. 2002 ¹²	Modified NEI-VFQ-25 plus supplement, Responsiveness	For each of 34 items, they compared the pre-treatment item difficulty to the post-item difficulty, and the difference was responsiveness. Item difficulty was based on the Rasch model. Thus, they measured whether certain visual tasks become easier after treatment.	7 of 34 items became statistically significantly easier after treatment (items 5, 6, 8, 14, A3, A4, and A8) (Figures 6 and 7a of the article). Item 5 is "How much difficulty do you have reading ordinary print in newspapers?". Item 6 is "How much difficulty do you have doing work or hobbies that require you to see well up close, such as cooking, sewing, fixing things around the house, or using hand tools?". Item 8 is "How much difficulty do you have reading street signs or the names of stores?". Item 14 is "Because of your eyesight, how much difficulty do you have going out to see movies, plays, or sports events?". Item A3 is "Wearing glasses, how much difficulty do you have reading the small print in a telephone book, on a medicine bottle, or on legal forms?" (item A3 was edited to include low vision devices as well as glasses). Item A4 is "Because of your eyesight, how much difficulty do you have figuring out whether bills you receive are accurate?". Item A8 is "Because of your eyesight, how much difficulty do you have seeing and enjoying programs on TV?"
Stelmack et al. 2002 ¹²	Modified NEI-VFQ-25 plus supplement, Responsiveness	For each of 77 patients, they compared the pre-treatment visual ability to the post-treatment ability, and the difference was responsiveness. Visual ability was based on the Rasch model. Thus, they measured whether certain patients became more able after treatment.	69 of 77 patients had a higher estimate of visual ability after treatment vs. before treatment (Figure 9 in the article). The typical amount of improvement corresponded to a 4-line improvement in visual acuity.
Stelmack et al. 2002 ¹²	Modified NEI-VFQ-25 plus supplement, Construct Validity	Authors used Rasch analysis to determine construct validity. Each item received a weighted fit statistic, and they determined whether the fit statistics before treatment were independent from fit statistics after treatment.	For item difficulty, the data demonstrate construct validity, as there was no relation between pre-intervention and post-intervention fit statistics (Figure 4a of the article). They found the same result for person ability estimates (Figure 4c of the article)

ADL=activities of daily living; ADLMS=activities of daily living mobility and safety; AMD=age-related macular degeneration; CoV=coefficient of variation; CR_{.95}=coefficient of reliability; dB=decibel; EWB=emotional well-being; IR=internal reliability; MD=macular dystrophy (I, II designate disease severity); NEI-VFQ-25=National Eye Institute Visual Function Questionnaire 25

Table C-9. Psychometric data reported in studies included for Key Question 1C (continued)

item; OR=other retinal dystrophy; PSI=person separation index; RP=retinitis pigmentosa (I-IV designate disease severity); SD=standard deviation; SST=Suprachoroidal Transretinal Stimulation device , VA-13=Veteran's Administration-13