**Table C-4. Psychometric data reported in studies included for Key Question 1B**

| Study | Measure and Psychometric Property | Description of Analysis | Result |
|---|---|---|---|
| Geruschat et al. 2015[1] | Functional Low-Vision Observer Rated Assessment (FLORA), ***Face validity*** | Initial item generation | A team of experts in blind and low vision rehabilitation met to draft a first assessment. Multiple rounds of revision in the suggested FLORA process. The team reviewed commonly-accepted instruments and tailored FLORA to the challenges of this population. Face validity is suggested by the fact that the participants were experts in this clinical area, as well as the various steps they undertook. |
| Geruschat et al. 2015[1] | Functional Low-Vision Observer Rated Assessment (FLORA), ***Face validity*** | Whether the self-report questions were used by most assessors | For 22/26 patients, all 14 questions were answered. In the other 4, an average of 12 questions were answered. |
| Geruschat et al. 2015[1] | Functional Low-Vision Observer Rated Assessment (FLORA), ***Face validity*** | Whether the part 2 activities were used by most assessors | The average number of patients assessed per activity was 20 out of a possible 26 (see Table 1 of the article). This indicates that assessors tended to ask patients to perform most of the FLORA activities of daily living. |
| Bittner et al. 2011[2] | Grating Acuity Test (GAT), ***Construct Validity*** | Separately for the 8 patients with RP and the 12 patients with OR, authors computed the correlation between the newly developed GAT and the standard and "well-validated" test, ETDRS. Perfect validity would be indicated by (1) strong correlation, (2) a slope of 1.0 and (3) an intercept of 0. | GAT demonstrated this type of construct validity for patients with RP, but not for patients with OR (Figure 2 in the article). The correlations, slopes, and intercepts were not reported. For RP, the correlation was strong, the slope was near 1.0, and the intercept was near 0. For OR, however, the correlation was weak, the slope was greater than 1, and the intercept was about 0.75. This means that for patients with OR, the newly developed GAT consistently overestimated patients' visual acuity. |
| Bittner et al. 2011[2] | Grating acuity test (GAT), ***Reliability*** | Test-retest reliability. Authors computed each patient's coefficient of reliability, $CR_{.95}$. This was done both within-visit and between-visit. A low $CR_{.95}$ indicates good test-retest reliability, since it indicates the degree of difference between 2 tests that one might expect (a test with perfect test-retest reliability would have a $CR_{.95}$ of 0). Data were on the log-unit scale. | For RP, the median test-retest $CR_{.95}$ of GAT was 0.17 for within-visit and 0.16 for between-visit (log-unit scale, see Figure 5). For OR, the median test-retest $CR_{.95}$ of GAT was 0.11 for within-visit and 0.11 for between-visit (log-unit scale, see Figure 5 of the article). |
| Bittner et al. 2011[2] | ETDRS visual acuity test, ***Reliability*** | Test-retest reliability, same as above | For RP, the median test-retest $CR_{.95}$ of ETDRS was 0.10 for between-visit (log-unit scale, see Figure 5 in the article). For OR patients, it was 0.16. |

**Table C-4. Psychometric data reported in studies included for Key Question 1B (continued)**

| Study | Measure and Psychometric Property | Description of Analysis | Result |
|---|---|---|---|
| Chow et al. 2010[3] | Grating Acuity Test (GAT), ***Construct Validity*** | Authors computed the correlation between the newly developed GAT and the standard and "well-validated" test, ETDRS. Perfect validity would be indicated by (1) strong correlation, (2) a slope of 1.0, and (3) an intercept of 0. | For RP specifically, correlation was strong (r=0.92), the slope estimate was 0.92, and intercept not reported (but appeared to be about 0.02 from Figure 35 in the article). Thus, good results for RP. For AMD and other retinopathies, however, GAT consistently underestimated logMAR (i.e., overestimated visual acuity). The mean logMAR for patients with AMD was 1.4 as measured by the gold standard ETDRS but was only 0.89 as measured by GAT. For OR, the mean logMAR as measured by ETDRS was 1.37 as compared to 0.98 for GAT. Thus, for non-RP patients, GAT has poor validity. |
| Chow et al. 2010[3] | Grating Acuity Test (GAT), ***Reliability*** | Test-retest reliability. Authors computed each patient's coefficient of reliability, $CR_{.95}$. This was done both within-visit and between-visit. A low $CR_{.95}$ indicates good test-retest reliability. Data were on the log-units scale. | The mean test-retest $CR_{.95}$ of GAT was 0.16 for between-visit. For RP patients specifically, the mean test-retest $CR_{.95}$ of GAT was 0.15 for between-visit and 0.10 for within-visit. |
| Chow et al. 2010[3] | Chow Color Test (CCT), ***Construct Validity*** | Authors computed the correlation between the newly developed CCT and the standard test called the PV-16. The tests are on different scales, so only the strength of correlation is a relevant measure for construct validity. Because higher scores on the CCT mean better color vision, whereas higher scores on the PV-16 mean worse color vision, good validity would be indicated by a large negative correlation. | The correlation between CCT and PV-16 was r=-0.77. Patients averaged 22.5 out of 40 on the CCT, and they averaged 315 on the PV-16. |
| Chow et al. 2010[3] | Chow Color Test (CCT), ***Reliability*** | Test-retest reliability. Authors computed each patient's coefficient of reliability, $CR_{.95}$. This was done only between-visit. A low $CR_{.95}$ indicates good test-retest reliability. Data were on the same scale as the CCT, which is 0 (lowest possible color vision) and 40 (best possible color vision). | The mean test-retest $CR_{.95}$ of CCT was 6.1 for between-visit. For the 5 patients with AMD, it was 3.9; for the 5 patients with RP, it was 4.8; for the other 7 patients it was 8.7. The 3 groups' mean scores on the CCT were AMD, 30; RP, 13; and other, 24. Thus for an average RP patient, if their color vision testing was at 13 out of 40 at one visit, then the next visit would be expected (with 95% confidence) to be between 8 and 18 out of 40. |
| Kiser et al. 2005[4] | ETDRS visual acuity test, regular, ***Reliability*** | Test-retest reliability. Authors computed each patient's coefficient of reliability, $CR_{.95}$. This was done only between-visit. A low $CR_{.95}$ indicates good test-retest reliability, because it indicates the degree of difference between 2 tests that one might expect (a test with perfect test-retest reliability would have a $CR_{.95}$ of 0). Data were on the log-unit scale. | Median values of $CR_{.95}$ were:<br>0.13 for RP-I, 0.23 for RP-II, 0.26 for RP-III<br>0.27 for MD-I, 0.21 for MD-II<br>0.18 for DR<br>0.20 for OR<br>See Figure 3 of the article |

**Table C-4. Psychometric data reported in studies included for Key Question 1B (continued)**

| Study | Measure and Psychometric Property | Description of Analysis | Result |
|---|---|---|---|
| Kiser et al. 2005[4] | ETDRS visual acuity test, dim, ***Reliability*** | Test-retest reliability, same as above | Median values of $CR_{.95}$ were:<br>0.12 for RP-I, 0.41 for RP-II, 0.18 for RP-III<br>0.33 for MD-I, 0.20 for MD-II<br>0.27 for DR<br>0.19 for OR<br>See Figure 3 of the article. |

AMD=age-related macular degeneration; DR=diabetic retinopathy; ETDRS=Early Treatment of Diabetic Retinopathy Study (test); logMAR=logarithm of the minimum angle of resolution**;** MD=macular degeneration (I, II indicate better to worse visual acuity); OR=other retinopathies; RP=retinitis pigmentosa (I, II, III indicate better to worse visual acuity)