Effect of Examiner Experience and Technique on the Alternate Cover Test

Heather A. Anderson[*], **Ruth E. Manny**[*], **Susan A. Cotter**[†], **G. Lynn Mitchell**[‡], and **Jasmine A. Irani**[§]

University of Houston College of Optometry, Houston, Texas (HAA, REM), Southern California College of Optometry, Fullerton, California (SAC, JAI), and The Ohio State University College of Optometry, Columbus, Ohio (GLM).

Abstract

Purpose—To compare the repeatability of the alternate cover test between experienced and inexperienced examiners and the effects of dissociation time and examiner bias.

Methods—Two sites each had an experienced examiner train 10 subjects (inexperienced examiners) to perform short and long dissociation time alternate cover test protocols at near. Each site conducted testing sessions with an examiner triad (experienced examiner and two inexperienced examiners) who were masked to each other's results. Each triad performed the alternate cover test on 24 patients using both dissociation protocols. In an attempt to introduce bias, each of the paired inexperienced examiners was given a different graph of phoria distribution for the general population. Analysis techniques that adjust for correlations introduced when multiple measurements are obtained on the same patient were used to investigate the effect of examiner and dissociation time on each outcome.

Results—The range of measured deviations spanned 27.5 prism diopters ( ) base-in to 17.5 base-out. The absolute mean difference between experienced and inexperienced examiners was 2.28 ± 2.4 and at least 60% of differences were 2 . Larger deviations were measured with the long dissociation protocol for both experienced and inexperienced examiners (mean difference range = 1.17 to 2.14 , $p < 0.0001$). The percentage of measured small deviations (2 base-out to 2 base-in) did not differ between inexperienced examiners biased with the narrow vs. wide theoretical distributions ($p = 0.41$). The magnitude and direction of the deviation had no effect on the size of the differences obtained with different examiners or dissociation times.

Conclusions—Although inexperienced examiners differed significantly from experienced examiners, most differences were <2 suggesting good reliability of inexperienced examiners' measurements. Examiner bias did not have a substantial effect on inexperienced examiner

Heather A. Anderson, College of Optometry, University of Houston, 505 J Davis Armistead Bldg, Houston, TX 77204-2020, handerson@optometry.uh.edu.
[*]OD, PhD, FAAO
[†]OD, MS, FAAO
[‡]MAS, FAAO
[§]BS

measurements; however, increased dissociation resulted in larger measured deviations for all examiners.

Keywords

alternate cover test; dissociation; ocular deviation; phoria; prism neutralization; repeatability

The alternate cover test is a commonly performed objective clinical test to measure the magnitude of a deviation in the alignment of the two eyes, either strabismic or phoric, and is performed at distance and near using an occluder and prisms (loose or prism bar) as the patient views a target to control fixation and accommodation. The patient is instructed to fixate the target while the examiner alternately covers each eye to disrupt fusion (termed dissociation) and then adds prism to quantify the deviation in prism diopters ( ). Previous studies conducted with nonstrabismic subjects investigated both intra- and interexaminer repeatability of the alternate cover test and found excellent repeatability with absolute mean differences <2 and 95% confidence intervals of 4 or less.[1,2] All these studies report excellent repeatability of the alternate cover test; however, this excellent repeatability may have been produced because experienced examiners who had developed and refined their skills over time were compared. Although the threshold for detection of small eye movements does not differ between inexperienced and experienced examiners,[3] the ability to detect an eye movement may not correlate directly with the ability to perform the alternate cover test accurately.

In addition to observational skills, one major factor that may affect measurements made with the alternate cover test could be the total time spent dissociating a patient and the care with which an examiner dissociates the patient. Studies of eye movements during an automated cover test have shown that it takes at least 5 s of occlusion for the full horizontal deviation to manifest.[4] This occlusion time must not be interrupted with any brief periods of binocular vision that could drive the vergence of the eyes back toward the near target rather than toward the phoric position.[5] Thus, if an examiner were to perform the alternate cover test hastily with fewer than 5 s of total dissociation time or if the examiner were to move the cover too slowly between the eyes and permit intermittent periods of binocular vision, the full magnitude of the deviation may not be manifest.

An additional factor that may affect alternate cover test measurements is individual examiner bias about the expected distribution of deviations in the general population. For example, if an inexperienced examiner were taught that the majority of subjects are orthophoric (zero deviation) then that examiner may measure a greater prevalence of orthophoria in a sample population due to his or her expectations. In such a case, the biased examiner might not use careful observation in an attempt to detect small eye movements other than orthophoria or might only perform a few brief covers and stop immediately after observing orthophoria rather than attempting to dissociate for a longer time, which might allow a larger deviation to manifest.

The overall aim of this study was to investigate the repeatability between inexperienced and experienced examiners using the alternate cover test at different test sites, as might be done in a multicenter clinical trial. Specifically, this study design compared two different alternate

cover test protocols (short vs. long dissociation time) and the effect of examiner bias on the repeatability of inexperienced and experienced examiner measurements.

As noted above, differences in experience, technique, and expectation or bias on the part of an examiner can all impact a measurement outcome and produce significant issues for multicenter trials. The aim of this study was to investigate how these three factors impact the objective measurement of a deviation by the alternate cover test by comparing the results of inexperienced examiners trained on two different dissociation protocols (short vs. long dissociation time), and shown different expected results, to that of an experienced examiner (gold standard) at two different sites.

METHODS

This study was funded through a cooperative agreement with the National Eye Institute of the National Institutes of Health and was conducted at the Southern California College of Optometry (SCCO) in Fullerton, CA, and at the University of Houston College of Optometry (UHCO) in Houston, TX. The protocol and informed consent forms were reviewed and approved by the respective Institutional Review Boards. All study subjects gave written informed consent.

Participants

Study subjects consisted of examiners and patients. The eligibility criterion for inexperienced examiners was persons aged 18 to 65 years who were naive to the alternate cover test. Thus, no eye care providers or optometry students served as inexperienced examiners. The subjects came from the surrounding communities and included friends and family, university students, and staff members. The eligibility criterion for patients was any person 18 to 40 years of age with 20/30 or better visual acuity in each eye at near. Study subjects were recruited from the college campuses and surrounding communities. In addition, there was an experienced pediatric optometrist at each site (SAC at SCCO and REM at UHCO, each with 26 or more years of experience) who served as the gold standard examiner.

Examiner Training

A training program for performing the alternate cover test was developed and provided by the gold standard examiners to all inexperienced examiners. In an attempt to standardize the alternate cover test, the training program consisted of a Powerpoint presentation given by the gold standard examiner that explained how to perform the alternate cover test. The presentation contained embedded video clips of alternate cover test measurements being conducted according to the study protocol on patients with orthophoria, esophoria, or exophoria. Two variations (long and short dissociation, both described below) of the alternate cover test protocol were taught during these training sessions. This presentation was followed by two supervised hands-on training sessions where the examiners practiced the alternate cover test on subjects with orthophoria and various magnitudes of esophoria and exophoria. During the second training session, each inexperienced examiner was required to demonstrate his or her proficiency with both alternate cover test protocols. No

information regarding the distribution of phorias in the general population was provided during these training sessions.

Alternate Cover Test Protocols

Monocular visual acuity was measured immediately before testing to confirm that the patients had 20/30 or better visual acuity at near for the testing session. The alternate cover test was performed at 40 cm with the patients wearing their habitual refractive corrections (contact lenses or spectacles). Examiners instructed the patients to look at a near target with an overall size of approximately 20/67 and multiple smaller details (the house on the Children's Fixation bar no. 7004001; Clement Clarke, Harlow, Essex, UK) and to "keep the house clear." A black plastic cover paddle and a loose prism set (increments of 1, 2, 4, 6, 8, 10, 12, 15, 20, 25, 30, 35, 40, 45, and 50) were used to conduct the alternate cover test.

Examiners first determined whether the patient was orthophoric, eso, or exo by performing either the long or short alternate cover test protocol (described below). To ensure that the patient was always fully dissociated during testing, the protocol specified that the examiner hold the cover paddle as close as possible to the patient's eye, keeping the eye covered at all times, both when alternating the cover paddle and when interposing prisms. Furthermore, the protocol specified that the alternate cover test be performed with the left eye fixating (i.e., holding the prism before the right eye and neutralizing the movement of the right eye). Care was taken not to tilt or rotate the prism. The prism neutralization endpoint was recorded as the high neutral finding (i.e., the highest prism power that induced no movement before reversal of the deviation). If a neutral endpoint was not demonstrated using the prism amounts available in the loose prism set, the endpoint was "bracketed" by averaging the largest prism power demonstrating the initial directional movement and the smallest prism power to cause reversal of movement. Any vertical deviation was ignored.

The number of alternations of the cover paddle, combined with the total time of dissociation before the judgment of the direction of the deviation, varied according to the assigned protocol. For the short dissociation protocol, the cover paddle was initially held over the patient's left eye for 1 s, after which each eye was alternately occluded for a count of 1 s for three more excursions of the cover paddle (covering the right and left eye two times each for a total dissociation time of 4 s) to determine whether the patient was ortho, eso, or exo. For the long dissociation protocol, the cover paddle was initially held over the patient's left eye for 5 s, after which each eye was alternately occluded for a count of 2 s for nine additional excursions of the cover paddle (covering the right and left eye five times each; total dissociation time of 23 s) to determine whether the patient was ortho, eso, or exo. For both protocols, if no movement was observed, the patient was labeled orthophoric and testing stopped. If an eso or exo deviation was observed, the examiner neutralized the deviation with the alternate cover test procedure described above. There were no stipulations regarding the number of times the patient was alternately occluded or the number of prisms that could be interposed to determine the endpoint for the alternate cover test.

Testing Sessions

An examiner triad consisting of the gold standard examiner and two inexperienced examiners (one with each bias as described below) performed the alternate cover test using the long and short protocols as described previously. There were 10 examiners at each clinical site and 45 and 52 patients at the SCCO and UHCO sites, respectively. Each examiner participated in eight testing sessions with each session consisting of an examiner triad testing three different patients. Thus, each inexperienced examiner performed the alternate cover test on 24 different patients. Patients participated in one to five testing sessions with testing schedules arranged, so that no patient was ever tested twice by the same inexperienced examiner; however, because the gold standard examiner participated in all sessions, each of the gold standard examiners tested some of the same patients at her site during more than one test session.

Each patient was seated comfortably in a clinic examination room while the examiners rotated among the rooms and performed a short and long alternate cover test (but not sequentially) on each of the patients. To control for order effects, both the order of examiner and the order of long vs. short dissociation were randomized. The testing order to be assigned to each patient was randomly selected from a list of possible combinations of three examiners and two dissociation protocols. Therefore, each patient had his or her near deviation measured six times (three examiners using both protocols separated in time). Patients were given at least a 3-minute break between each alternate cover test measurement.

A few minutes before data collection, the gold standard examiner independently (without the other examiners' knowledge) presented each inexperienced examiner with his or her data collection sheet on which there was a printed phoria distribution, purportedly for the general population, on the second page. One inexperienced examiner was given the narrow theoretical distribution (Fig. 1A) and the other inexperienced examiner given the wide theoretical distribution (Fig. 1B). Both distributions were centered at zero (orthophoria); however, the narrow distribution indicated almost 90% orthophoria in the general population, whereas the wide theoretical distribution indicated <30% orthophoria. The intent of providing each examiner with a different distribution was to provide that person bias in his or her expectations of near alternate cover test measurements.

Statistical Analysis

A mixed linear model was used to assess the relationship between bias, study site, and dissociation protocol on the percent of alternate cover test measurements between 2 base-out (BO) and 2 base-in (BI) among the patients measured by each inexperienced examiner. This range of deviations was selected for analysis because 2 has been demonstrated to be the minimal magnitude eye movement detectable by an examiner and thus anything greater may represent a clinically significant finding.[3,6] To compare measurements with the short vs. long dissociation protocols, the absolute value of the difference between the measurements (long minus short) was evaluated within each examiner for each patient. Absolute differences were calculated rather than signed differences due to the nature of the

data (exo is negative, and eso is positive) to avoid artificially minimizing the differences through positive and negative differences cancelling each other out.

For comparisons of dissociation protocol, the greatest difference in total dissociation time would occur when orthophoria was recorded for the short protocol (4 s dissociation time for the short vs. a minimum of 23 s for the long). A mixed linear model was used to assess the relationship between examiner, study site, and short dissociation result (orthophoria or not) on the absolute value of the difference between long and short dissociation. This same statistical technique was used to assess the relationship between bias, study site, and dissociation protocol on the absolute value of the difference between the inexperienced and experienced examiners' alternate cover test measurement. Mixed modeling was performed, so that the inherent correlation introduced by repeated measurements on the same patient could be controlled. A mixed linear model was also used to estimate the variability as required for constructing the Bland-Altman 95% limits of agreement.[7] General Estimating Equation models were used to evaluate the effect of examiner (inexperienced or experienced), bias, and site on the classification of an alternate cover test measurement between 2 BO and 2 BI. All statistical analyses were performed using SAS (Version 9.2, SAS institute, Cary, NC).

RESULTS

Measurement Distribution

The range of alternate cover test measurements in this study spanned 27.5 BI to 17.5 BO (Fig. 1A to D). Fig. 1A and B shows the distribution of alternate cover test measurements using the short vs. long dissociation protocols for inexperienced examiners biased with the narrow theoretical distribution (1A) and those biased with the wide theoretical distribution (1B). Fig. 1C and D shows the distribution of alternate cover test measurements for inexperienced vs. experienced examiners for the short dissociation protocol (1C) and long dissociation protocol (1D).

Effects of Examiner Bias and Dissociation Protocol on Measurement Distribution

Table 1 shows the percentage of alternate cover test measurements between 2 BO and 2 BI for each of the examiner groups using each protocol. For the inexperienced examiners, the percentage of alternate cover test measurements between 2 BO and 2 BI was not related to the theoretical distribution used to bias each group's findings (p = 0.41). There was also no significant relationship between proportion of alternate cover test measurements between 2 BO and 2 BI and site (p = 0.31) and no significant interactions between site and the other covariates (bias and protocol, p > 0.10). The distribution of alternate cover test measurements obtained by the inexperienced examiners was significantly related to measurement protocol. A smaller percentage of measurements fell between 2 BO and 2 BI when the long dissociation protocol was used by both groups of inexperienced examiners (p < 0.001).

The distribution of measurements for the experienced examiners was a smoother plot than that of the inexperienced examiners, which had multiple peaks (Fig. 1C and D). It was

hypothesized that the peaks in the inexperienced examiners distributions corresponded to the increments of loose prisms available for neutralization of the deviation, suggesting that inexperienced examiners did not interpolate measurements as often as experienced examiners. A *post hoc* analysis confirmed that inexperienced examiners were significantly less likely to interpolate their alternate cover test measurements (Fig. 1C and D), when compared with the experienced examiner at each study site (UHCO $p < 0.001$, SCCO $p = 0.044$).

Comparison of Dissociation Protocols

The effect of examiner, study site, and short dissociation result (orthophoria or not) on the absolute difference between the long and short dissociation protocols was examined because the total dissociation time with the short protocol differs when orthophoria is observed vs. using prism to neutralize a deviation. There was no significant relationship with study site ($p = 0.30$), and there were no significant interactions between site and the other covariates ($p > 0.10$). Table 2 shows the absolute mean differences of alternate cover test measurements for the long and short dissociation protocols both overall and also by short dissociation result and examiner. In all cases, the absolute mean difference between dissociation protocols was significantly greater than zero ($p < 0.0001$), indicating that larger deviations were measured with the long dissociation protocol. The absolute mean differences were also found to vary significantly with the result of the measurement obtained with the short dissociation protocol (orthophoria or not) and the examiner obtaining the measurement (interaction $p = 0.003$). When measurements were made by the experienced examiners, there was a significantly greater difference between long and short dissociation values if the result of the short dissociation was orthophoria (mean difference = 2.14) rather than nonorthophoria (mean difference = 1.44 , $p = 0.004$). The magnitude of the difference between short and long dissociation protocols were not related to the results of the short dissociation measurement for inexperienced examiners biased with the narrow ($p = 0.26$) or wide ($p = 0.19$) population distributions. Fig. 2A is a difference vs. mean plot[7] comparing the two dissociation protocols and demonstrates that the magnitude of the measurement differences was unrelated to the magnitude of the mean deviation.

Comparison of Inexperienced vs. Experienced Examiners

Table 3 shows the absolute mean differences of alternate cover test measurements between the inexperienced and experienced examiners. The overall mean absolute difference was 2.28 ± 2.40 . The differences between examiners did not vary significantly with study site ($p = 0.059$); however, they differed with the observed distribution bias of the inexperienced examiners. Differences between examiners exposed to the wide distribution, and the experienced examiners were larger than differences between examiners exposed to the narrow distribution and the experienced examiners; however, the effect was small ($p = 0.046$).

The absolute differences in alternate cover test measurements between examiners varied significantly with dissociation protocol ($p = 0.010$) with greater differences between experienced and inexperienced examiners found with the long dissociation protocol (absolute mean difference = 2.44) vs. the short dissociation protocol (2.12). A difference

vs. mean plot[7] shown in Fig. 2B demonstrates that the differences between inexperienced and experienced examiners are not related to the magnitude of the mean deviation.

Comparison of the Percentage of Small-Angle Deviations Measured

Analyses were performed to assess the effect of bias, dissociation protocol, and study site on the agreement between the experienced and inexperienced examiners in classification of small angle deviations (2Δ BO to 2Δ BI). Given the observed differences in the alternate cover test measurements with long and short dissociation, the analyses were performed separately for each measurement protocol. Agreement between experienced and inexperienced examiners did not differ between the two study sites for the long or short dissociation protocols (p = 0.65, p = 0.15, respectively) nor were there significant interactions between site and the other covariates (p > 0.10). Table 4 shows a comparison of the percentage of alternate cover test measurements between 2Δ BO and 2Δ BI for experienced and inexperienced examiners using the long protocol and indicates that the inexperienced examiners were more likely to have measured a deviation between 2Δ BO and 2Δ BI (p = 0.004); however, there were no significant differences between the experienced and inexperienced examiners for the short dissociation protocol (Table 5, p = 0.14).

DISCUSSION

This study investigated the effect of examiner experience on measurements made with the alternate cover test and found a range of absolute mean differences from 1.91 to 2.56Δ with 95% confidence intervals of ±4.1 to ±5.3Δ around the mean (calculated as 1.96 × standard deviation from Table 3) when comparing experienced and inexperienced examiners. Some of these mean differences are just greater than the level of minimal detectable eye movement by an examiner (2Δ),[3,6] suggesting that they could be considered clinically significant. The differences found in this study are also greater than those reported by a previous study comparing alternate cover test measurements of experienced examiners (mean absolute differences = 1.19 to 1.67Δ with 95% confidence intervals less than ±4Δ).[2] Thus, alternate cover test measurements are less similar comparing inexperienced examiners to experienced examiners than comparing two experienced examiners. It should also be noted that for all testing conditions, at least 60% of differences in measurements between experienced and inexperienced examiners were not >2Δ, suggesting that more often than not the measurements of the inexperienced examiners were in good agreement with the experienced examiners and that for a minority of the measurements (40%) differences may be clinically significant.

Although the group absolute mean differences between experienced and inexperienced examiners were relatively small, there was a large range of individual differences in measurements (0 to 17.5Δ) (Table 3), with approximately 3% of differences > 10Δ. One potential explanation for the large magnitude differences could be recording errors by inexperienced examiners. For example, an inexperienced examiner who correctly neutralized an 8Δ exo deviation but recorded 8Δ BO instead of BI would have an absolute difference of 16Δ from the findings of an experienced examiner. Although this large difference is not representative of the inexperienced examiner's ability to perform the

physical aspects of the alternate cover test, it could highlight a potential problem of accurately documenting findings in a patient record. Interexaminer comparisons were investigated to look for these potential recording errors. Fifteen incidences of measurements differing by >10 were observed, but only three of these consisted of measurements of similar magnitude and opposite sign, indicating that the majority of large magnitude differences between experienced and inexperienced examiners were probably not related to recording errors.

In this study, inexperienced examiners were shown either a narrow theoretical distribution or a wide theoretical distribution of expected phoria findings in an attempt to bias the inexperienced examiners and to determine whether examiner bias impacts the distribution of alternate cover test measurements. As shown in Table 1, exposure to the narrow vs. wide theoretical distributions did not affect the distribution of the alternate cover test measurements in any substantial way. This could indicate that examiner bias does not have a significant impact on measurements with alternate cover test, or it could be that the technique used in this study was not sufficient to introduce a bias. After the study was completed, some of the inexperienced examiners anecdotally reported that they did not give much attention to the distribution they were shown, suggesting that a more stringent effort to bias the inexperienced examiners may have resulted in a different outcome. Future experiments with a more stringent method of introducing bias may be valuable for addressing this question.

Another purpose of this study was to determine the effect of dissociation time on alternate cover test measurements. As would be predicted from previous studies of time to break fusion,[4,5] this study found a significant increase in the magnitude of the deviation measured with the long dissociation protocol vs. the short dissociation protocol. One limitation of this study is that the total length of dissociation with the short protocol increased if the examiner detected movement and proceeded with neutralizing prism and additional alternate covers, potentially making it more similar to the long protocol. Given the above limitation in comparing dissociation time, one scenario was investigated in which the total difference in dissociation time between short and long protocols would be maximized. In the case of reporting orthophoria for the short protocol, the total dissociation time would have been limited to 4 s in the event that the examiner switched the cover paddle every second as instructed. The long dissociation protocol, in contrast, would consist of a minimum of nine alternate covers for a total dissociation time of at least 23 s (5 s initial cover and 2 s for each subsequent cover). The difference in measurement with the short vs. long protocol was greatest for experienced examiners when orthophoria was recorded for the short protocol (as shown in Table 2 and Fig. 1C and D). This finding is consistent with greater measured deviations obtained with prolonged dissociation time across examiner groups and suggests that an examiner who uses four, quick, 1 s covers to perform the alternate cover test would be more likely to report orthophoria, when in fact a larger deviation may be present. Although the absolute mean difference between short and long dissociation was only 2.14 for experienced examiners, the range of differences was quite large (up to 12), indicating that some large magnitude deviations (particularly eso deviations) would not have been observed if only a brief cover test had been performed.

Site differences between UHCO and SCCO were not found in this study for the short vs. long dissociation protocols. Although the subjects at each study site were not randomly selected from the general population, subjects were also not selected specifically by the examiners but rather volunteered in response to email and flyer solicitations. It is not expected that subjects with a particular magnitude of deviation would be more likely to volunteer for the study than others, and thus it is predicted that the distribution of deviations between sites should not differ from that of the general population. The lack of differences between sites is encouraging, and an important finding for multicenter clinical studies in which the alternate cover test is used to measure near deviations. This study suggests that using a standardized method with a well-defined dissociation protocol will reduce or eliminate differences between study sites.

One potential limitation to this study is that some patients participated in more than one session, and thus the experienced examiners would have tested a particular patient multiple times (for a maximum of five different sessions) and potentially been biased about their expected findings for the repeated patients. However, each experienced examiner performed 240 cover tests throughout the completion of the study with sessions being separated by several hours to several days. Thus, it is unlikely that the experienced examiners would recall the previous cover test findings of any one particular patient.

Another potential limitation to this study is that subjects were not screened with a unilateral cover test at the time of enrollment to identify the presence or absence of strabismus, and thus, the sample of subjects in this study may have included both subjects with and without strabismus; however, the vast majority of subjects in this study were nonstrabismic. Previous studies have found the repeatability of the alternate cover test decreases for examiners testing strabismic subjects.[8,9] It is possible that the differences between experienced and inexperienced examiners would have been greater if a larger percentage of subjects with strabismus were included in the study. In addition, all the subjects in this study were cooperative adults, and it is possible the results may have differed had young children been included as cover test patients. It should also be noted that the results of this study should not be generalized to an inexperienced examiner's ability to identify the presence or absence of strabismus with the unilateral cover test.

CONCLUSIONS

The findings of this study indicate that binocular deviation measurements with the alternate cover test differ significantly between inexperienced and experienced examiners; however, the majority of differences (60%) were not clinically significant (<2) suggesting that inexperienced examiners' measurements are often reliable. The attempt at creating examiner bias about the expected distribution of phorias did not significantly impact the distribution of alternate cover test measurements in this study. These findings suggest that inexperienced examiners can be trained to accurately perform the alternate cover test, given a well-defined protocol and specific training. This study also found larger deviations measured with longer dissociation time for both experienced and inexperienced examiners and a higher percentage of orthophoria with brief periods of dissociation (4 s). The magnitude and direction of the deviation had no effect on the size of the measured differences between examiners or

between dissociation protocols. In light of these findings, studies using multiple examiners and sites should use a rigid alternate cover test protocol and standardized examiner training with a defined minimum dissociation time of >4 s to maximize similarity between examiners and sites. These recommendations may also be applicable to faculty instructing new students or private practitioners training personnel to maximize similarity between examiners performing the alternate cover test.

Acknowledgments

REFERENCES

1. Rainey BB, Schroeder TL, Goss DA, Grosvenor TP. Inter-examiner repeatability of heterophoria tests. Optom Vis Sci. 1998; 75:719–726. [PubMed: 9798211]

2. Johns HA, Manny RE, Fern K, Hu YS. The intraexaminer and interexaminer repeatability of the alternate cover test using different prism neutralization endpoints. Optom Vis Sci. 2004; 81:939–946. [PubMed: 15592119]

3. Fogt N, Baughman BJ, Good G. The effect of experience on the detection of small eye movements. Optom Vis Sci. 2000; 77:670–674. [PubMed: 11147737]

4. Barnard NA, Thomson WD. A quantitative analysis of eye movements during the cover test—a preliminary report. Ophthalmic Physiol Opt. 1995; 15:413–419. [PubMed: 8524567]

5. Fogt N, Toole AJ. The effect of saccades and brief fusional stimuli on phoria adaptation. Optom Vis Sci. 2001; 78:815–824. [PubMed: 11763255]

6. Ludvigh E. Amount of eye movement objectively perceptible to the unaided eye. Am J Ophthalmol. 1949; 32:649. [PubMed: 18121093]

7. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986; 1:307–310. [PubMed: 2868172]

8. Pediatric Eye Disease Investigator Group. Interobserver reliability of the prism and alternate cover test in children with esotropia. Arch Ophthalmol. 2009; 127:59–65. [PubMed: 19139339]

9. Holmes JM, Leske DA, Hohberger GG. Defining real change in prism-cover test measurements. Am J Ophthalmol. 2008; 145:381–385. [PubMed: 18045567]
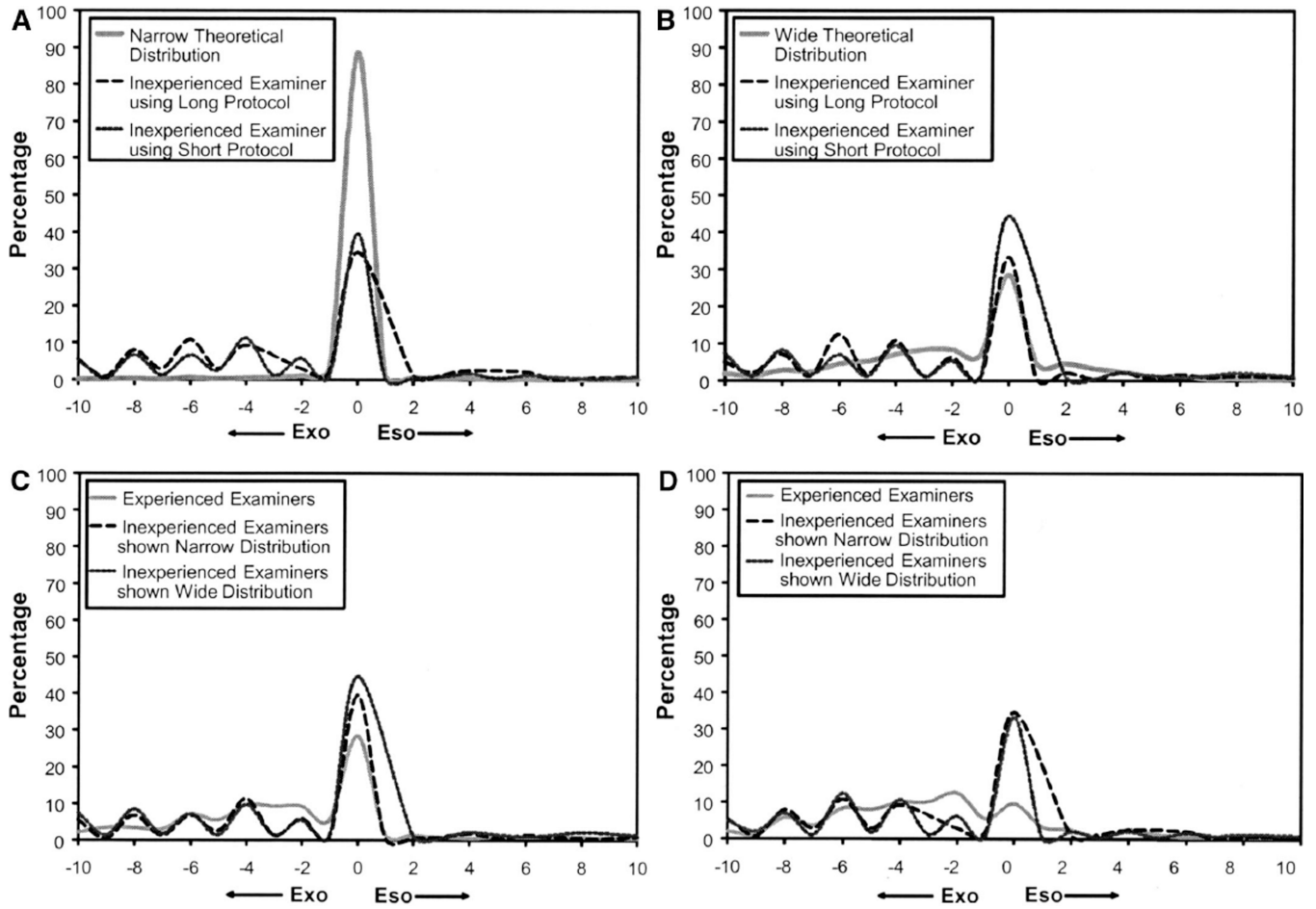
FIGURE 1.
Distribution of alternate cover test measurements for inexperienced examiners shown the narrow theoretical distribution (A) vs. the wide theoretical distribution (B). Distribution of alternate cover test measurements for inexperienced vs. experienced examiners using the short dissociation protocol (C) vs. the long dissociation protocol (D).
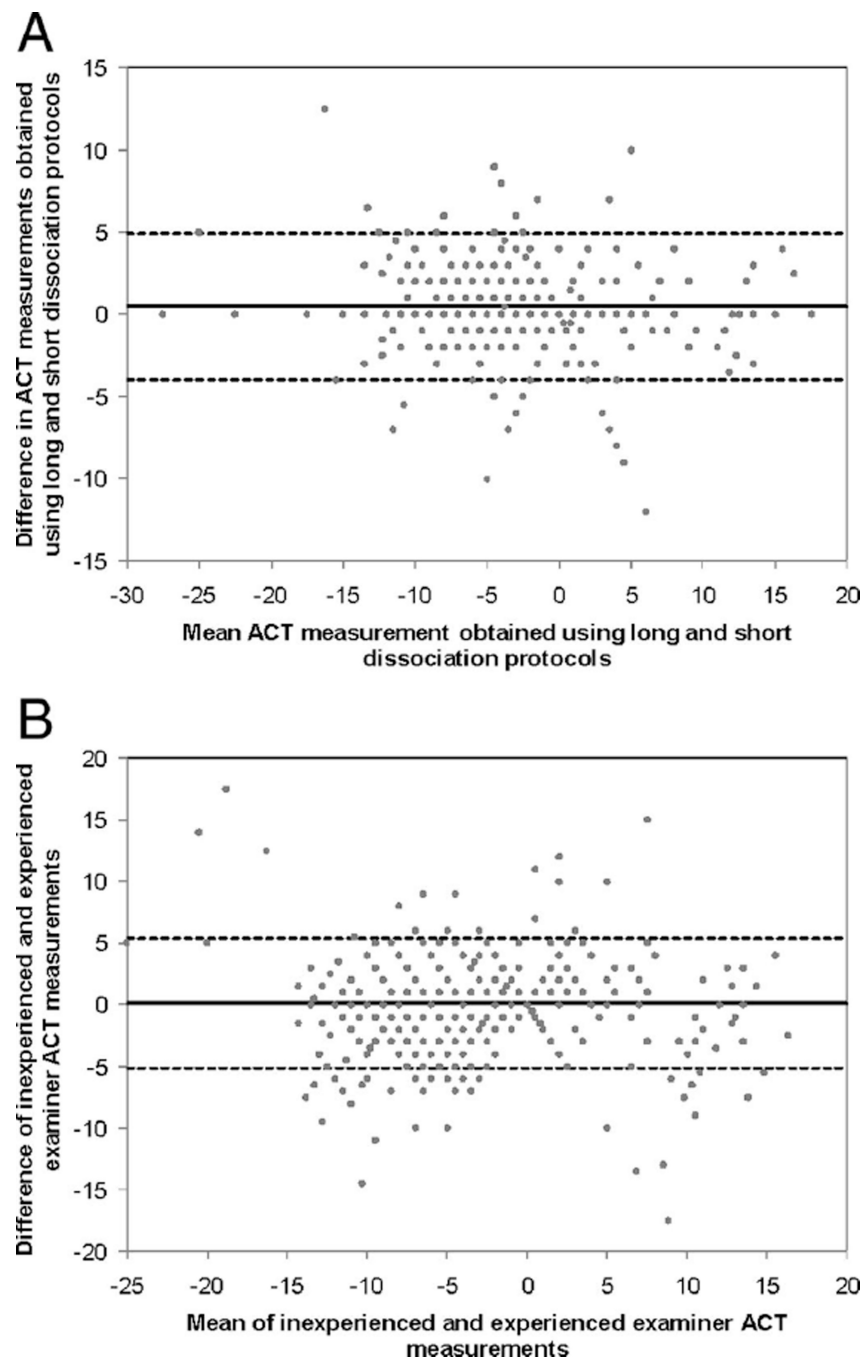
FIGURE 2.
Difference vs. mean plot for the short vs. long dissociation protocols (A) (mean difference = 0.50 , 95% limits of agreement: −3.95 , +4.95) and difference vs. mean plot for experienced vs. inexperienced examiners (B) (mean difference = 0.16 , 95% limits of agreement: −5.10 , +5.42).

TABLE 1

Distribution of alternate cover test measurements

	Mean (Δ)	Median (Δ)	Range	Percentage between 2 Δ BO and 2 Δ BI
Narrow theoretical distribution	0	0		95.0
Inexperienced examiners shown the narrow distribution when using the long protocol	2.9 BI	2 BI	22 BI–15 BO	40.8
Inexperienced examiners shown the narrow distribution when using the short protocol	2.5 BI	Ortho	22 BI–17 BO	47.1[a]
Wide theoretical distribution	0	1 BO		53.0
Inexperienced examiners shown the wide distribution when using the long protocol	3.1 BI	2 BI	22 BI–15 BO	42.9
Inexperienced examiners shown the wide distribution when using the short protocol	2.6 BI	Ortho	17 BI–17 BO	52.1[a]
Experienced examiner				
Using the long protocol	3.2 BI	3 BI	27.5 BI–17.5 BO	33.3
Using the short protocol	2.7 BI	2 BI	27.5 BI–17.5 BO	45.8[a]

[a]Significantly more deviations between 2 Δ BO and 2 Δ BI were measured with the short dissociation protocol for all examiner groups (p < 0.001).

TABLE 2

Comparison of alternate cover test measurements with long vs. short dissociation protocol

	Absolute mean (SD) difference ( )	Range ( )
All examiners combined		
Ortho w/short dissociation (n = 268)	1.49 (2.3)[a]	0–12
Nonortho w/short dissociation (n = 452)	1.56 (1.7)[a]	0–12.5
Overall	1.53 (1.9)[a]	0–12.5
Experienced examiners		
Ortho w/short dissociation (n = 67)	2.14 (2.1)[a]	0–12
Nonortho w/short dissociation (n = 173)	1.44 (1.4)[a]	0–7
Overall	1.64 (1.7)[a]	0–12
Inexperienced examiners shown narrow distribution		
Ortho w/short dissociation (n = 94)	1.17 (2.3)[a]	0–9
Nonortho w/short dissociation (n = 146)	1.57 (1.6)[a]	0–7
Overall	1.41 (1.9)[a]	0–9
Inexperienced examiners shown wide distribution		
Ortho w/short dissociation (n = 107)	1.36 (2.2)[a]	0–8
Nonortho w/short dissociation (n = 133)	1.71 (2.1)[a]	0–12.5
Overall	1.55 (2.2)[a]	0–12.5

Differences are the absolute value of the measurement by long dissociation minus the measurement by short dissociation. "n" indicates number of measurements.

[a] All mean differences were significantly greater than zero (p < 0.0001).

TABLE 3

Comparison of alternate cover test measurements by inexperienced vs. experienced examiners

	Absolute mean (SD) difference ( )	Range ( )	Percentage differences >2
Overall	2.28 (2.4)	0–17.5	35.9
Long protocol			
Shown narrow distribution	2.31 (2.1)[a]	0–14.5	38.8
Shown wide distribution	2.56 (2.5)[a]	0–17.5	39.2
Overall	2.44 (2.3)[a]	0–17.5	39.0
Short protocol			
Shown narrow distribution	1.91 (2.3)	0–15	30.8
Shown wide distribution	2.34 (2.7)	0–17.5	35.0
Overall	2.12 (2.5)	0–17.5	32.9

Differences are the absolute value of the measurement difference of the inexperienced examiner minus the experienced examiner.

[a] Differences were significantly larger when examiners used the long dissociation protocol vs. the short dissociation protocol (p = 0.01).

TABLE 4

Agreement between inexperienced and experienced examiners using the long dissociation protocol

	Measures between 2△BO and 2△BI?	Inexperienced examiner	
		Yes	No
Experienced examiner	Yes	114 (23.8%)	46 (9.6%)
	No	87 (18.1%)[a]	233 (48.5%)

[a] Inexperienced examiners were more likely to record a measurement between 2△BO and 2△BI (p = 0.004).

TABLE 5

Agreement between inexperienced and experienced examiners using the short dissociation protocol

	Measures between 2Δ BO and 2Δ BI?	Inexperienced examiner	
		Yes	No
Experienced examiner	Yes	177 (36.9%)	43 (9.0%)
	No	61 (12.7%)	199 (41.5%)

Categorization did not differ between examiners when using the short dissociation protocol (p = 0.14).