SPECIAL ARTICLE

A report of the ABO Resident Clinical Outcome Study (the pilot study)

Vance J. Dykhouse,[a] Allen H. Moffitt,[b] John E. Grubb,[c] Peter M. Greco,[d] Jeryl D. English,[d] Barry S. Briss,[d] Scott A. Jamieson,[d] Marvin C. Kastrop,[d] and S. Ed Owens, Jr[e]

Blue Springs, Mo, Murray, Ky, Chula Vista, Calif, Philadelphia, Pa, Houston, Tex, Boston, Mass, Marquette, Mich, Billings, Mont, and Jackson, Wyo

Introduction: A 4-year collaborative project between the American Board of Orthodontics (ABO) and 15 American orthodontic graduate programs concluded at the ABO Clinical Examination in February 2006. **Methods:** Fifty recent graduates (the pilot study group) presented cases that were treated in their graduate programs as partial fulfillment of the requirements for ABO certification. The examinees were tested by calibrated ABO examiners and required to satisfy traditional ABO standards. They presented appropriate case reports that contained the ABO's 3 measurement instruments. Thirty-nine practicing orthodontists (the control group) presented cases according to the traditional ABO Clinical Examination process. **Results:** Ninety percent of the pilot study group and 85% of the control group successfully accomplished ABO certification. There was a difference of 2.38 points between the 2 groups for the mean total scores of the cases that passed. The pilot study group presented cases that met the historic averages for case complexity (discrepancy index). **Conclusions:** Residents in orthodontic programs are able to treat cases to ABO standards of quality. (Am J Orthod Dentofacial Orthop 2006;130:656-61)

As early as the autumn of 2001, the American Board of Orthodontics (ABO) discussed the possibility of certifying orthodontists after their orthodontic graduate training education. Other health professions,[1,2] including dental specialties,[3-5] have instituted early certification pathways after residency completion. Integral and essential to the early certification concept would be the requirement of periodic recertification throughout the practitioner's career.

The ABO directors decided to study whether orthodontic residents could treat to ABO standards. A protocol was devised, and orthodontic programs were contacted to enlist their collaboration on the project. The ABO included a cross-section of orthodontic programs from the separate American Association of Orthodontists constituencies comprising diverse program lengths and class sizes in both public and private institutions. By the summer of 2002, 20 programs had tentatively agreed to participate. Ultimately, 15 programs signed protocol agreements, and their former residents presented cases at the ABO Clinical Examination in February 2006.

This article will review the ABO Resident Clinical Outcomes Study (the pilot study), which demonstrated that orthodontic residents can present cases that meet the same certification standards required of conventional ABO examinees.

METHODS

The participating orthodontic program directors and the incoming residents for 2002 were asked to select 12 cases for prospective treatment and inclusion as the pilot study (PS) cases. The contractual protocol included the following stipulation: "Cases should be representative of a cross section of clinical problems and of adequate difficulty to represent the resident's ability to diagnose and treat orthodontic patients needed to fulfill the Commission on Dental Accreditation (CODA) Standards #4-3.1, #4-3.2, and #4-3.3."[6] The cases must have been treated by only 1 resident under the direct supervision of clinical faculty.

Copies of pretreatment intraoral and cephalometric radiographs were sent to the ABO for each of the 12 cases selected by the residents. These would eventually be used to verify that each PS case was an original prospectively selected case. The protocol requested submission of up to 6 finished cases satisfying the ABO case display requirements. Prospective treatment for the resident's cases was a strict requirement to parallel

[a]President of the ABO.
[b]President-elect of the ABO.
[c]Secretary-treasurer of the ABO.
[d]Director of the ABO.
[e]Immediate past president of the ABO.
Reprint requests to: Dr Vance J. Dykhouse, American Board of Orthodontics, 401 N Lindbergh Blvd, Suite 308, St Louis, MO 63141; e-mail, info@americanboardortho.com.
Submitted, August 2006; revised and accepted, September 2006.
0889-5406/$32.00

doi:10.1016/j.ajodo.2006.09.002

the ABO's previously established Option II certification pathway.

In October 2004, the ABO directors significantly altered the certification process as announced in the March 2005 issue of *American Journal of Orthodontics and Dentofacial Orthopedics*.[7] An integral part of the new process was the Initial Certification Examination intended for graduates of approved programs of the American Dental Association, Council on Dental Accreditation (CODA) who had successfully completed the ABO's written examination. The PS investigation was refocused. Not only would it examine whether residents could treat cases to ABO standards during their specialty programs, but also it would evaluate realistic case criteria for future resident case presentations.

The PS residents were allowed to present cases for the ABO Clinical Examination (formerly the Phase III examination). This stipulation was that these cases must include at least 1 PS case but preferably up to 6 PS cases. Additionally, the residents would be allowed to present up to 6 supplemental cases with a minimum discrepancy index (DI) of 7. The supplemental case differed from the PS case in that the supplemental case was not prospective. A total of 6 cases, including the supplemental cases, would be required to meet ABO standards to earn a 10-year time-limited certificate. A 15-year time-limited certificate would be issued if a candidate completed all 6 PS cases (without supplemental cases) in an attempt to encourage residents to present the full requirement of 6 prospective PS cases.

This expansion of the project allowed the ABO to evaluate a number of factors beyond the original PS objectives. The PS participants (PS group) prepared all cases to satisfy ABO case display requirements, which included use of the same measurement instruments as in the traditional certification examination including the DI,[8] the cast evaluation form,[9] and the case management form[10] for each case. These instruments provided additional numerical data for the study. Also, a summary sheet for each PS participant was completed by the ABO examiner, which documented each case by treatment length, categorized the treatment plan by extraction or nonextraction therapy, and provided a record of whether the case satisfied ABO standards.

Finally, a brief postexamination survey of the ABO examiners was conducted by requesting them to complete a numerical rating scale to compare the quality of the oral responses of the resident examinees with those of traditional candidates.

The ABO examiners (ABO directors and invited consultants) were all calibrated before the Clinical Examination in February 2006. The PS group and the control group (2006 regular ABO examinees) were both tested by the same 20 calibrated examiners. The customary ABO examination protocol was followed for each group to negate differences in examiners who were calibrated at different times. The examinees were randomly assigned to the examiners. The examiners were unaware of the examinees' identities until immediately before the oral examinations. Examiners could not examine a candidate with whom they had a previous affiliation.

The control group prepared their case displays and were asked to complete all 3 measurement instruments for their cases. As with the PS participants, these self-scored measurements were verified by the ABO examiner and corrected by the examiners as needed. The sample populations for both groups were thus randomly obtained by self-selection via their agreement to participate in the ABO Clinical Examination.

The statistical analyses performed by using the studies' data consisted of descriptive statistics (means, medians, standard deviations, and ranges) for the PS cases, the supplemental cases from the PS group, and the control group cases. The chi-square test,[11] the Fisher exact test,[11] the linear mixed model,[12] and the *t* test[13] were all used as appropriate to compare the PS and the control groups.

RESULTS

The 50 PS participants presented a total of 422 cases, of which 243 (58%) were from the prospectively selected PS cases and 179 (42%) were supplemental cases. The average number of total cases presented by the PS group was 8.44 per examinee. The control group's 39 examinees presented 326 cases, or an average of 8.40 cases per examinee.

Table I summarizes the statistics of both groups. The PS cases and the supplemental cases had similar descriptive statistics. Comparison of the PS (293) with the supplemental cases (179) shows that the average pass rates were 84% for the PS cases and 81.6% for the supplemental cases, but these differences are not statistically significant (Table II). Therefore, the supplemental cases were not included in these data in comparisons with the control group because the supplemental cases did not significantly influence the data. Also, the supplemental cases were not included in the PS's original protocol.

Further comparison of the complete and incomplete PS cases shows that the mean total scores (cast evaluation score plus case management form score) were 21.70 for the complete cases and 37.67 for the incomplete cases. This 16-point difference was statistically significant at $P < .001$ (Table III). There were only 39

Table I. Statistical summaries of raw data

	Total	*Ratio*	*DI*	*CE*	*CMF*
Pilot cases (n = 243)					
Mean	24.26	0.77	16.96	21.98	3.18
SD	8.08	0.51	9.28	6.5	2.74
Range	7-63	0-3.86	0-46	7-53	0-14
Median	23	0.65	15	20	3
Supplemental cases (n = 179)					
Mean	23.6	0.83	17.89	20.75	2.85
SD	7.62	0.52	10.33	6.65	2.61
Range	8-46	.15-2.61	4-68	5-41	0-16
Median	23	0.69	15	20	2
Control group (n = 326)					
Mean	20.21	1.26	21.84	17.44	2.77
SD	6.62	1.11	11.01	5.61	2.21
Range	3-41	.27-10.67	6-76	0-36	0-11
Median	20.5	1	20	17	2.5

Total, CE and CMF; *Ratio,* DI/total; *DI,* discrepancy index[8]; *CE,* cast evaluation form[9]; *CMF,* case management form.[10]

Table II. Comparison of pilot and supplemental cases

	Pilot	*Supplemental*
Total cases	243	179
Average cases per examinee	4.86	3.58
Range per examinee	2-6	1-6
Average pass rate*	84%	81.60%

*Number of passed cases divided by each participant's total cases presented for pilot and supplemental, respectively. Each participant received pass rate for both pilot and supplemental cases. Pass/fail rate was not significantly different between pilot and supplemental cases. Chi-square = 0.415; df = 1; $P = .520$.

Table III. Pilot cases passed or incomplete

Total score	*Passed (n = 204)*	*Incomplete (n = 39)*
Mean	21.7	37.67
SD	6.06	7.69
Range	7-32	24-63
Median	22	38

incomplete cases, but, when these were included in the summary of PS case mean values, the mean was elevated from 21.70 for complete cases to 24.26 for all cases.

The comparison of the PS group with the control group demonstrated that the percentage of pass (complete) cases and incomplete cases was not significantly different. Forty-five (90%) of the PS participants successfully completed the examination, and 5 (10%) were incomplete, in contrast to 33 (85%) successful 2006 traditional examinees and 6 (15%) incomplete. The Fisher exact test of $P = 0.525$ showed no significant difference between the groups.

Review of the summary of the base data of the 3 measurement instruments and the total score and ratio showed that the scores of the PS cases were significantly below those of the control group (Table IV). The mean total score for the PS cases was 24.26, whereas the control group's mean was 20.21. These numbers have a P value of $<.001$ that indicated a significant difference between the groups, with a difference of 4 points. However, both mean scores are passing for the Clinical Examination, and both the PS and the control groups met ABO standards. Only the case management scores were not statistically different, but, because of the low, single-digit number for those scores, it would be difficult to differentiate between them.

It is interesting to compare successfully completed PS cases with successfully completed control cases (Table V). The mean total scores were 21.70 for the PS cases and 19.32 for the control cases. Both total scores were reduced by elimination of the incomplete cases from the calculations. The mean total score for passing PS cases was reduced by 2.90 points from the mean total score from all cases (24.26), whereas the mean total score for the control cases (20.21) was reduced by only 0.89 points. The average difference in total scores between the 2 groups was 2.38 points compared with an average difference of 4 points for all cases considered, although the difference was still statistically significant ($P = .005$). This value indicates that the scores for the incomplete cases of the PS group influenced the overall mean to a greater degree than those of the control group. When comparing the passed cases, the 2 groups were closer on the average total scores. This is meaningful because all passed cases meet ABO standards.

The length of active treatment for all PS cases was 24.67 months (SD, 4.80). For all cases presented, both PS and supplemental cases, 244 (59%) were nonextraction, and 173 (41%) were extraction.

The survey responses of the 20 examiners who participated in the 2006 Clinical Examination were interesting. The survey was intended to solicit each examiner's subjective evaluation of the PS participants' oral examinations compared with the examiner's previous experience with traditional ABO examinees. Fifty-nine percent of the examinees were judged by the ABO examiners to have understood and discussed their cases with equal skill levels to those of traditional ABO examinees. Sixty-four percent of the examinees were judged by the ABO examiners to have understood the ABO's diagnosis and treatment planning cases with skill levels equal to those of traditional ABO examinees.

Table IV. Pilot vs control group cases

	Pilot cases (n = 243)	*Control cases (n = 326)*	P *value**
Total scores			
Mean/SD	24.26/8.08	20.21/6.62	*P* <.001
Range	7-63	3-41	
Ratio	0.77/0.52	1.26/1.12	*P* <.001
DI			
Mean/SD	16.96/9.28	21.84/11.01	*P* <.001
Range	0-46	6-76	
Cast evaluation			
Mean/SD	21.08/6.5	17.44/5.61	*P* <.001
Range	7-53	0-36	
Case management form			
Mean/SD	3.18/2.74	2.77/2.21	*P* = .21
Range	0-14	0-11	

*The Pilot Study cases and the Control cases were compared using multilevel mixed models. The test statistics and the degrees of freedom are available if the reader is interested.

Table V. Passed cases only: pilot vs control group

	Pilot cases (n = 204)	*Control cases (n = 274)*	P *value**
Total scores			
Mean/SD	21.70/5.06	19.32/6.12	*P* = .005
Range	7-32	3-31	
Ratio	0.83/0.53	1.29/1.18	*P* = .001
DI			
Mean/SD	17.02/9.27	20.99/10.2	*P* <.001
Range	0-46	6-73	
Cast evaluation			
Mean/SD	19.16/4.23	16.75/5.27	*P* = .001
Range	7-27	0-7	
Case management form			
Mean/SD	2.54/2.05	2.57/1.99	*P* = .997
Range	0-7	0-7	

*The Pilot Study cases and the Control cases were compared using multilevel mixed models. The test statistics and the degrees of freedom are available if the reader is interested.

DISCUSSION

The traditional ABO examination process is an outcomes assessment of an orthodontist's knowledge and clinical abilities after his or her orthodontic training and years of clinical experience. A number of previous studies used various indexes to evaluate the quality of orthodontic therapy.[14-18] Several articles specifically explored the results of treatment performed in an academic environment.[19-21]

This study was unique. It was a randomized collection of orthodontists' best cases from across the United States judged by examiners who were calibrated for specific measurement instruments (indexes) with respect to historic ABO quality standards. The quality of treatment was correlated with the venue in which the cases were treated. The cases of the PS group were compared with those treated in orthodontists' private offices (control group). The results are meaningful as an outcomes measure not only for the orthodontic educational system but also for practicing orthodontists in the United States. Both groups performed admirably.

Eighty-five percent of the practicing orthodontists were successful in their certification endeavors. Ninety percent of recent orthodontic-program graduates were also successful. Eighty-four percent of the PS group cases and 88% of the control group cases passed the examinations. These statistics should be a source of pride for the entire specialty.

In February 2005, the ABO published a study that compared the DI of the prospectively selected PS cases with the mean DI of the 2003 and 2004 traditional ABO examinees' cases.[22] The PS prospectively treated cases (n = 857) had a mean DI of 17.20, and the traditional ABO examinees' cases (n = 625) had a DI of 16.10. In

contrast, our study's control group cases (n = 326) had a mean DI of 21.84 compared with the PS cases (n = 243) DI of 16.96 (Table IV). The 2 DI scores from this study were significantly different at $P < .001$. It should be obvious by comparing the DI mean from the 2 studies that the 2006 traditional examinees (control group) had a higher mean DI than their predecessors. The PS participants brought cases to the 2006 Clinical Examination that had a mean DI (16.96) almost identical to the mean DI (17.20) of the entire prospectively selected case population in the February 2005 report. The conclusion can therefore be made that the PS group cases had reasonable case complexity compared with both the control group and previous ABO examination cases.

The therapeutic outcome for each case can be measured by the total score (cast evaluation score plus the case management form score). Comparison of the mean total score of the PS cases (24.26) to the mean total score of the control group cases (20.21) showed a significant difference at $P < .001$ (Table IV). Both total score means would satisfy ABO standards. The control group, representing experienced orthodontists, was only 4 points better on average than the recently graduated PS group. This difference was certainly expected and even welcomed. One would hope that greater clinical experience would improve treatment quality. Yet the 4-point difference decreased to a 2.38-point difference when only the successfully completed cases from each group were compared (means, 21.70 for completed cases from the PS group and 19.32 from the control group; Table V).

Even though there was a statistically significant difference between the control group and the PS group, the differences were small. When the cast evaluation scores for the 2003 and 2004 ABO Clinical Examination (688 cases with a mean score of 15.47) were compared with the 2006 Control group (326 cases with a mean score of 17.44), there was only a 1.97-point difference. Statistically, this 2-point difference in the mean scores is still significantly different at $P = .009$. This would mean a significant difference between the mean CE scores for the traditional ABO examinees between various years of the examinations. Thus, when evaluating the statistics, the reader must understand the relationship between sample size and statistical power. The larger the sample size, other things being equal, the higher the power of the test, and the higher the probability of detecting significance.[23]

The difference in mean total scores for the passed cases between the PS group (large sample size of 204 cases) and the Control group (large sample size of 274 cases) can be considered small at 2.38 points. Because these cases all satisfy ABO standards, the summary statistics are interpreted to mean that orthodontic residents are capable of treating to ABO expectations.

An additional favorable observation of the PS group is the subjective opinions of the ABO examiners regarding the PS participants' performance on their oral examinations compared with traditional ABO examinees. Fifty-nine percent of the PS participants were perceived to perform as well in discussion of their presented cases as past examinees. Sixty-four of the PS participants were thought to perform as well as past examinees when discussing their examination cases. Most ABO examiners were pleasantly surprised by the overall knowledge and abilities of the young orthodontists as shown in examination discussion.

No comparisons were made between the 2 groups as to numbers of extraction cases or treatment durations. There was no prerequisite for the number of extraction cases required for display. The data for the PS group demonstrated that 41% of their cases involved extraction of teeth. The mean treatment time for the 244 PS cases was 24.67 months (SD, 4.8 months; range, 7-40 months); this compares favorably with previous studies on duration of orthodontic treatment.[24,25]

CONCLUSIONS

The ABO collaborated with 15 orthodontic programs and established the Resident Clinical Outcomes Study to determine whether orthodontic residents can treat to ABO standards. This 4-year project concluded in February 2006 and provided an affirmative answer to that question. The results indicate that orthodontic residents can treat to ABO standards within the time frame of graduate orthodontic programs. The residents can present reasonably complex cases for examination, effectively treat these cases in acceptable treatment times, and discuss those cases with sufficient knowledge of the diagnoses and therapies performed.

This information will be used by the ABO to establish fair and appropriate criteria for the Initial Certification Examination. The ABO believes that the successful results of this extensive project provide positive affirmation of the high quality of education provided by our orthodontic educational system.

We thank the ABO central office staff and especially Prudy Yerkes for recording the data; Lingling Zhang, statistician with the Center for Statistical Consultation and Research, University of Michigan, for expert statistical analysis and advice; and Allen Moffitt for his organizational and administrative efforts during the project.

We also extend our gratitude to the chairpersons,

directors, faculty, and graduates of the following programs that participated in this project: Albert Einstein Medical Center, Baylor College of Dentistry, Case Western Reserve University, Harvard School of Dental Medicine, Mayo Graduate School of Medicine, University of Illinois at Chicago, University of Kentucky, University of Maryland, University of Michigan, University of North Carolina, University of Tennessee, University of Southern California, University of Texas at Houston, University of Texas at San Antonio, and West Virginia University.

REFERENCES

1. American Board of Pediatrics. Available at: www.abp.org. Accessed on June 14, 2006.
2. American Board of Internal Medicine. Available at: www.obim.org. Accessed on June 14, 2006.
3. American Board of Pediatric Dentistry. Available at: www.abpd.org. Accessed on June 14, 2006.
4. American Board of Oral and Maxillofacial Pathology. Available at: www.aaomp.org. Accessed on June 14, 2006.
5. American Board of Oral and Maxillofacial Surgeons. Available at: www.aboms.org. Accessed on June 14, 2006.
6. Advanced Specialty Education Programs in Orthodontics and Dental Facial Orthopedics. Commission on Dental Accreditation, American Dental Association. January 2000, standards 4-3.1, 4-3.2, and 4-3.3.
7. Riolo ML, Owens SE, Dykhouse VJ, Moffitt AH, Grubb JE, Greco PM, et al. A change in the certification process by the American Board of Orthodontics. Am J Orthod Dentofacial Orthop 2005;127:278-81.
8. Cangiolosi TJ, Riolo ML, Owens SE, Dykhouse VJ, Moffitt AH, Grubb JE, et al. The ABO discrepancy index: a measure of case complexity. Am J Orthod Dentofacial Orthop 2004;125:270-8.
9. Casko JS, Vaden JL, Kokich VG, Damone J, James RD, Cangiolosi TJ, et al. Objective grading system for dental casts and panoramic radiographs. Am J Orthod Dentofacial Orthop 1998;114:589-99.
10. Owens SE, Dykhouse VJ, Moffitt AH, Grubb JE, Greco PM, English JD, et al. The case management form of the American Board of Orthodontics. Am J Orthod Dentofacial Orthop 2006; 129:325-9.
11. Agresti A. Categorical data analysis. 2nd ed. New York: Wiley-Interscience; 2002.
12. Snijders T, Bosker R. Multilevel analysis: an introduction to basic and advanced multilevel modeling. Thousand Oaks, Calif: Sage; 1999.
13. Utts J, Heckard R. Mind on statistics. Belmont, Calif: Thomson-Brooks/Cole; 2004.
14. Richmond S, Shaw WC, Roberts CT, Anderson M. The PAR index (Peer Assessment Rating): methods to determine outcome of orthodontic treatment in terms of improvement and standards. Eur J Orthod 1992;14:180-7.
15. Summers CJ. The occlusal index: a system for identifying and scoring occlusal disorders. Am J Orthod 1971;59:552-66.
16. Pickering EA, Vig P. The occlusal index used to assess orthodontic treatment. Br J Orthod 1975;2:47-51.
17. Holman JK, Hans MG, Nelson S, Powers MP. An assessment of extraction versus non-extraction orthodontic treatment using the peer assessment rating (PAR) Index. Angle Orthod 1998;68: 527-34.
18. Templeton KM, Powell R, Moore MB, Williams AC, Sandy JR. Are the peer assessment rating index and the index of treatment complexity, outcome and need suitable measures for orthognathic outcomes? Eur J Orthod 2006; e-pub April 28.
19. Firestone AR, Hasler RU, Ingervall B. Treatment results in dental school orthodontic patients in 1983 and 1993. Angle Orthod 1999;69:19-26.
20. Onyeaso CO, BeGole EA. Orthodontic treatment—improvement and standards using the peer assessment rating index. Angle Orthod 2006;76:260-4.
21. Pinskaya YB, Hsieh TJ, Roberts WE, Hartsfield JK. Comprehensive clinical evaluation as an outcome assessment for a graduate orthodontics program. Am J Orthod Dentofacial Orthop 2004; 126:533-43.
22. Riolo ML, Owens SE, Dykhouse VJ, Moffitt AH, Grubb JE, Greco PM, et al. ABO resident clinical outcomes study: case complexity as measured by the discrepancy index. Am J Orthod Dentofacial Orthop 2005;127:161-3.
23. Cohen J. Statistical power analysis for the behavioral sciences. New York: Academic Press; 1969. p. 6-8.
24. Vig PS, Weintraub JA, Brown C, Kowaiski CJ. The duration of orthodontic treatment with and without extractions: a pilot study of five selected practices. Am J Orthod Dentofacial Orthop 1990;97:45-51.
25. Fink DF, Smith RJ. The duration of orthodontic treatment. Am J Orthod Dentofacial Orthop 1992;102:45-51.