Meeting of the Alternatives to the Exam Development Committee
Oregon State Board of Bar Examiners
Wednesday, September 21, 2022
Zoom Meeting – Invites are sent via Outlook Calendar
Open Session Agenda
(Items may not be discussed in the order listed or may be discussed in a workgroup session during the meeting)

Wednesday, September 21, 2022, 12:00 p.m. – 1:45 p.m.

1. **Call to Order/Finalization of Agenda**

   A. Roll of Attendees

   B. Finalize Agenda

2. **Old Business/Updates on Events/News/Developments of Interest**

   A. Discussion of Schedule

   B. Lead BBX Member will report to meeting. After all reports, workgroups will report to their assigned Zoom:

   (Assigned Zoom Link is in Parenthesis) (Password and other login information below Zoom Link)

   i. Outreach Group (Stay in the Main Zoom Meeting Room, as this will convert to the Outreach Meeting) (No additional Password or Meeting ID needed)

      a. Discuss possible changes to messaging/Slides from last presentation

      b. What will be discussed in the next workgroup session

   ii. SPP (https://us02web.zoom.us/j/86073137875?pwd=Uko2V2xGd0FjMjgxMEFxem0vNVM0UT09)
   (Meeting ID: 860 7313 7875 - Passcode: 810235)

      a. What was discussed at the last workgroup session

      b. What are the goals for this upcoming workgroup session **See Exhibit 1**

      c. Is the timeline still achievable?

   iii. OEP (https://us02web.zoom.us/j/83299242180?pwd=QzRMaVRNbkhTODVrVmg3VXNYMjYwQT09)
   (Meeting ID: 832 9924 2180 - Passcode: 724456)

      a. What was discussed at the last workgroup session

      b. What are the goals for this upcoming workgroup session **See Exhibit 2**

      c. Is the timeline still achievable?

3. **New Business**

   A. Break into workgroup sessions

4. **Adjourn**

EXHIBIT 1

Comments from Joanne Kane for SPP Chapters 6 - 8

(A) Other new lawyers working for the Employer commonly perform those tasks;

(B) The Provisional Licensee is compensated at their regular wage for time spent on those tasks; and

(C) Time spent on these tasks constitutes no more than 10 percent of time spent on all tasks assigned to the Provisional Licensee.

5.5 Roles and Duties of Mentors. Mentors participating in the Program have the same role and duties as other Mentors appointed under the New Lawyer Mentoring Program.

Section 6
Program Requirements

6.1 Overview of Program Requirements. The Program has ten substantive requirements and an hours requirement, each described in more detail in the Rules below. The substantive requirements are:

(A) Diligent, competent, and professional work on all Legal Work assigned to the Provisional Licensee by their Supervising Attorney;

(B) Completion of the Professional Liability Fund's "Learning the Ropes" CLE program;

(C) Completion of the New Lawyer Mentoring Program in the manner described in Rule 6.4;

(D) Production of at least 8 independently authored pieces of written work product;

(E) Leadership of at least 2 initial client interviews or client counseling sessions;

(F) Leadership of at least 2 negotiations;

(G) Reflections on each of the 8 pieces of written work product, 2 client interviews or counseling sessions, and 2 negotiations;

(H) Development and maintenance of a learning plan for accomplishing the above activities;

(I) Completion of regular timesheets recording all time devoted to the Program; and

(J) A Portfolio organizing the above Program components.

6.2 Legal Work. Provisional Licensees will perform Legal Work assigned to them by their Supervising Attorney.

(A) Legal Work, as defined by Rule 1.2(I), includes any work that is commonly performed by licensed attorneys in Oregon. Legal work may include activities that are also performed by unlicensed individuals, as long as newly licensed attorneys regularly incorporate those activities in their work.

(B) Provisional Licensees must perform this work diligently, competently, and professionally.

(C) A Provisional Licensee should not attempt work for which they feel unprepared or incompetent to perform. Instead, the Provisional Licensee should discuss their reservations with the Supervising Attorney and seek appropriate assistance.

Commented [JK1]: Some previous research suggests very general terms like these could be open to interpretation and possibly shift as a function of individual difference and/or implicit bias. Can these be further defined/operationalized to help avoid those concerns?

Commented [JK2]: For D--F, is it a problem from a fairness perspective if the work varies significantly across participants in terms of complexity/difficulty? If not, why not?

Commented [JK3]: Does it need to be "original" work? (Is it OK for example if one person drafts a brief pretty much from scratch and another person submits a simple will they created using a template and/or software?) Will/can the comparison still be apples-to-apples?

Commented [JK4]: I think we're all on the same page in terms of being supportive of formative opportunities to learn and reflect on learning. My sense is that those should that be somewhat separate from the materials submitted specifically for purposes of licensure, though, given that licensure is typically a more summative assessment.

Commented [JK5]: (See above. Will there be further definitions of what this looks like and doesn't? Training?)

(D) Provisional Licensees may request particular types of Legal Work that would benefit their professional development or completion of this Program. Provisional Licensees may also request permission to work on pro bono matters. If a Supervising Attorney agrees to any of these requests, that agreement constitutes "assignment" of Legal Work.

6.3 Learning the Ropes. The Provisional Licensee must attend or watch all sessions of the 2021 or 2022 "Learning the Ropes" program offered by the Oregon State Bar Professional Liability Fund. When the Provisional Licensee has satisfied this requirement, they must include their certificate of completion in their Portfolio.

6.4 New Lawyer Mentoring Program (NLMP). The NLMP Manual explains all required and optional elements of the NLMP.

(A) Provisional Licensees must follow the directions in the Manual for completing the required activities in Parts 1 to 5; at least one optional activity in each of those Parts; and at least 10 practice activities.

(B) Provisional Licensees must use the Manual worksheets to document their completion of required activities, optional activities, and practice activities. Those worksheets must be included in each Quarterly Portfolio, updated to reflect all activities completed to date.

(C) Provisional Licensees may not use any of the required activities specified in Rules 6.5 to 6.11 to satisfy the required activities, optional activities, or practice activities required by the NLMP. All NLMP activities are in addition to the other requirements of this Program.

(D) Lawyers who obtain their permanent licenses through the Provisional Licensing Program will not have to repeat the NLMP during their first year of practice. That requirement will be waived for lawyers completing the NLMP as part of the Provisional Licensing Program.

6.5 Independently Authored Written Work Product.

(A) Each Provisional Licensee must prepare and submit at least 8 pieces of independently authored work product. Written work product may take any form that lawyers use in their practices: memos, letters, emails, complaints, motions, briefs, contracts, wills, etc. All submitted work product, however, must comply with the following requirements:

(1) The work product must address some *substantive* aspect of a legal matter. Provisional Licensees may not submit routine scheduling notes or engagement letters for this component of the Portfolio.

(2) At least 2 of the pieces of work product must exceed 1500 words, not including headers or signature blocks. Footnotes do count towards the word total.

(3) All submitted work product must be *independently* authored by the Provisional Licensee. If another individual edits the Provisional Licensee's work, the Provisional Licensee must submit the original draft without those edits.

Commented [JK6]: Might suggest including a maximum and minimum length on all of them to give participants the clearest possible understanding of what will qualify. (Emails can be pretty short, for example; might not give a lot of information/be very useful in the evaluation.)

If Provisional Licensees are unsure whether a particular piece of work product qualifies for submission, they may ask one or both Program Managers. The Program Managers will not judge the quality of the work product, but can advise the Provisional Licensee whether the work product meets these minimum requirements.

(B) Each piece of work product must be accompanied by a brief assessment completed by the Supervising Attorney. The Board will provide standardized rubrics to use for these assessments.

(C) Each piece of work product must also be accompanied by the Provisional Licensee's structured reflection. The Board will provide a standardized template for that reflection.

(D) If the Provisional Licensee used a sample, template, or other work as foundation for the work product, the Provisional Licensee must also:

(1) Submit a copy of the original sample, template, or other work used as a foundation; and

(2) Highlight the portions of the work product that represent the Provisional Licensee's additions, edits, or other customization.

(E) If the work product relates to a client matter, and the work product has not been publicly filed in substantially the same form:

(1) The work product must be redacted to protect the client's interests; and

(2) The client must consent to inclusion of the work product in the Portfolio.

(F) If the Provisional Licensee is unable to gather sufficient work product from client-related work, the Supervising Attorney may assign a mock exercise or exercises to the Provisional Licensee, which can be submitted to fulfill this requirement. All mock work product must comply with subsections (A) through (D) above.

6.6 Client Interviews or Counseling Sessions. Each Provisional Licensee must lead two client interviews or counseling sessions that are observed and assessed by their Supervising Attorney.

(A) Requirements for the Provisional Licensee:

(1) Before beginning each interview or counseling session, the Provisional Licensee must explain the Supervising Attorney's role, assure the client that the Supervising Attorney (a member of the Provisional Licensee's organization) is encompassed by the attorney-client privilege, and obtain the client's consent to the Supervising Attorney's presence.

(2) If the client objects to the Supervising Attorney's presence, the Supervising Attorney will depart, and the Provisional Licensee will conduct the session without assessment. The Provisional Licensee must find another interview or counseling session to satisfy this component of the Program.

(3) After completing the interview or counseling session and receiving feedback from the Supervising Attorney, the Provisional Licensee must complete a structured reflection about the interview or counseling session. The Board will provide a template for that structured reflection.

Commented [JK7]: Again, this is sounding a bit more appropriate for a formative assessment than a summative one.

On the one hand, it might seem like it wouldn't "hurt" to have this, and given it is useful for candidates to have formative/reflective opportunities, it seems like it'd be fine/good to include. But on the other hand, it might be hard to predict how the Examiners would (consciously or not?) use this. If a candidate did not perform sufficiently well for a "pass," but can say why they underperformed and perhaps commit to change, is that "good enough" for purposes of the licensure decision? (If so, might need to adjust "claim"/IUA slightly. If not, why ask for it?)

Commented [JK8]: What if very little has been changed/customized? Is that OK? If not, may need to add some additional rules.

Commented [JK9]: (Would it be the Supervising Attorney's responsibility to ensure this had been properly done?)

Commented [JK10]: Sounds great to me! If you had a little library of mock exercises, it could both offer opportunities for Supervisors and Participants to see examples of what level of complexity/length would be appropriate and to substitute in work if the participant didn't have enough that would qualify.

As always, would just have to be careful/attuned to fairness issues associated with item/exercise reuse.

Commented [JK11]: Consider whether collecting a transcript could be an option?

Commented [JK12]: Again, sounds like a great formative opportunity. Wondering whether Examiners would see this and what the pros/cons would be.

(B) Role of the Supervising Attorney:

(1) The Supervising Attorney should not speak during the interviews or counseling sessions other than to greet the client at the beginning of each session, provide input specifically requested by the Provisional Licensee, or intervene if necessary to protect the interests of the client.

(2) If the client addresses a question to the Supervising Attorney, the Supervising Attorney should politely redirect the question to the Provisional Attorney.

(3) For each of the two interviews or counseling sessions, the Supervising Attorney will complete the client interview/counseling session rubric provided by the Board, share the completed rubric with the Provisional Licensee, and offer any additional feedback that would assist the Provisional Licensee's development.

(4) The Supervising Attorney's completed rubric will become part of the Provisional Licensee's Portfolio.

Commented [JK13]: If the Supervisor ended up substantively contributing, would the whole session be ineligible?

Commented [JK14]: It'd be helpful to better understand what the role and main judgmental task of the Examiner from the BBX will be. Is it mostly just verifying that all the steps were completed/functionally "agreeing" with the Supervising Attorney's assessment? Is it to independently assess the candidate's portfolio of work against an external criterion held constant across candidates? If the former, makes sense to include the Supervisor's Assessments. If more the latter, might not want to include the supervisor's assessments to avoid "biasing" the examiner.

(C) Exceptions:

(1) Provisional Licensees working in a prosecutor's office may use interviews or other sessions with a complainant to satisfy this component of the Program. The provisions in subsections (A) and (B) will govern those sessions except that the attorney-client privilege will not apply.

(2) If a Provisional Licensee does not conduct client (or complainant) interviews or counseling sessions in their workplace, the Program Managers will help the Provisional Licensee find other ways to satisfy this component of the Program. It may be possible, for example, for the Provisional Licensee to interview simulated clients as part of a law school competition or CLE exercise. Any substitutes will be structured to follow the other portions of this Rule as closely as possible.

6.7 Negotiations. Each Provisional Licensee must conduct two negotiations that are observed and assessed by the Supervising Attorney. A negotiation includes any discussion aimed at reaching an agreement. It may occur in the context of litigation, transactional, regulatory, or other matters. The negotiation does not have to focus on final resolution of the matter; it may focus on preliminary or interim matters.

(A) Requirements for the Provisional Licensee:

(1) Before beginning each negotiation, the Provisional Licensee must explain the Supervising Attorney's role to other attorneys participating in the negotiation. If the client is present, the Provisional Licensee must also explain the Supervising Attorney's role to the client and obtain the client's consent to the Supervising Attorney's presence.

(2) If the client objects to the Supervising Attorney's presence, the Supervising Attorney will depart, and the Provisional Licensee will conduct the negotiation without that assessment. The Provisional Licensee must find another negotiation to satisfy this component of the Program.

Commented [JK15]: Again, could a transcript be possible?

(3) After completing the negotiation and receiving feedback from the Supervising Attorney, the Provisional Licensee will complete a structured reflection about the negotiation. The Board will provide a template for that structured reflection.

Commented [JK16]: (See prev. comments.)

(B) Role of the Supervising Attorney:

(1) The Supervising Attorney should not speak during the negotiations other than to greet other attorneys and the client at the beginning of each session, provide input specifically requested by the Provisional Licensee, or intervene if necessary to protect the interests of the client.

(2) If another attorney or the client addresses a question to the Supervising Attorney, the Supervising Attorney should politely redirect the question to the Provisional Attorney.

(3) For each of the two negotiations, the Supervising Attorney will complete the negotiation rubric provided by the Board, share the completed rubric with the Provisional Licensee, and offer any additional feedback that would assist the Provisional Licensee's development. The completed rubric will become part of the Provisional Licensee's Portfolio.

(C) Exceptions: If a Provisional Licensee does not conduct negotiations in their workplace, the Program Managers will help the Provisional Licensee find other ways to satisfy this component of the Program. It may be possible, for example, for the Provisional Licensee to conduct negotiations as part of a law school competition or CLE exercise. Any substitutes will be structured to follow the other portions of this Rule as closely as possible.

6.8 Reflections. The Provisional Licensee must complete reflections, using templates provided by the Board, for (a) each of the 8 independently authored written work products submitted under Rule 6.5; (b) each of the 2 client interviews or counseling sessions conducted under Rule 6.6; and (c) each of the 2 negotiations conducted under Rule 6.7. The Provisional Licensee must include all of these reflections in their Portfolio.

Commented [JK17]: See previous comments. These seem incredibly helpful for the provisional licensee to learn/develop/grow as a professional. So I can definitely see the appeal of this from a formative/educational perspective. Seem less appropriate in a summative context like a licensure decision, which is generally less about improvement/development/change and more about readiness.

6.9 Learning Plan. The Learning Plan has two purposes: (a) It tracks what the Provisional Licensee has accomplished to date, and (b) it sets goals for the coming quarter.

(A) The Board will provide a Learning Plan template to help Provisional Licensees create and update their Learning Plans.

Commented [JK18]: A general one, or customized? Might have to see examples to fully understand what the Learning Plans will be. Progress monitoring sounds potentially resource-intensive.

(B) Provisional Licensees must update the Learning Plan regularly and include the current Plan in each Quarterly Portfolio.

(C) The final Learning Plan, submitted as part of the Final Portfolio, will demonstrate how the Provisional Licensee has accomplished all requirements of the Program. It will also require the Provisional Licensee to reflect on how they will continue to improve their skills and knowledge as a practicing lawyer.

Commented [JK19]: Echoing several previous comments, this sounds very helpful in general but also very different from most summative assessments I'm familiar with.

6.10 Timesheets. Provisional Licensees must complete timesheets for every day spent working in the Program. These timesheets are separate from any timekeeping required by the

Supervising Attorney or Employer. The Program timesheets are used to document achievement of the 1500 hours required by the Program and to demonstrate the Provisional Licensee's competence in timekeeping.

(A) The Board will provide timesheet blanks for Provisional Licensees to use for these purposes.

(B) Provisional Licensees may use those timesheets to record their time in either 6- or 15-minute increments, whichever is most convenient for them.

(C) Decimals must be used to record all time. E.g., 30 minutes is .5 hours.

(D) Provisional Licensees should record all time devoted to the Program, even if that time is not billable to a client.

(E) Timesheets must be submitted as part of both Quarterly Portfolios and the Final Portfolio.

6.11 Portfolio. The Provisional Licensee must create and maintain a Portfolio collecting all of the above materials. The Board will provide a template for organizing the Portfolio. As explained in Section 7 below, Provisional Licensees must submit their Portfolio for review at the end of every quarter they participate in the Program. When the Provisional Licensee is ready to seek admission to the Oregon State Bar, the Provisional Licensee must prepare and submit a Final Portfolio as provided in Rule 9.2.

Commented [JK20]: Consider pros/cons/expenses/scalability associated with having the BBX only review the Final Portfolio as opposed to quarterly/interim portfolios? Clarify the role of the BBX/Examiner vs. the Supervisor? (Should we explicitly note somewhere that the BBX Examiner and Supervisor must be independent/two different people?)

6.12 Hours. To qualify for admission to the Oregon State Bar, Provisional Licensees must document at least 1500 hours spent working within the Provisional Licensing Program. Those hours may include:

(A) All time spent devoted to Legal Work assigned by the Supervising Attorney, even if the time is not billed to a client;

(B) All time devoted to working on the Program components outlined in Rules 6.3 to 6.11.

Commented [JK21]: Just clarifying that the exceptions would be included? If I am going to include a classroom activity, I could retroactively apply the classroom credits/time? If I retroactively apply a competition, is it just the time I was competing? Time preparing? Will it feel unfair to candidates if different individuals count the time differently? If that's a concern, can more examples/clarity be given about how to account for hours?

(C) All time spent discussing or learning about Program components with their Supervising Attorney, other Employees of the Employer, their Mentor, the Program Managers, or an Ombudsperson;

(D) All time spent reviewing or reflecting on feedback from the Board on Quarterly Portfolios;

Commented [JK22]: ?

(E) All time spent in any training or educational activities required by their Employer that are not included in the Program components; and

(F) Up to 30 additional hours of MCLE activities.

Section 7
Quarterly Portfolios and Review

7.1 Requirement of Quarterly Portfolios. Provisional Licensees must submit their Portfolios to the Board each quarter that they participate in the Program. These Quarterly Portfolios must contain all Program work described by Rules 6.3 to 6.10 that has been completed to date.

Commented [JK23]: Again just wondering about the pros/cons/burdens/expense/scalability? Better to have BBX review only the final portfolio and have the more formative/progress-related assessments rest with (trained) supervisors?

7.2 Notification of Quarterly Due Dates. Regulatory Counsel will notify Provisional Licensees of the time and date when each Quarterly Portfolio is due, as well as the process for submitting the Quarterly Portfolio. That notice will be provided at least 30 calendar days before the due date.

7.3 Timeliness of Quarterly Portfolios. Lawyers frequently must comply with deadlines, and missing a deadline can have serious consequences for a client. Provisional Licensees, therefore, must submit at least 2 Quarterly Portfolios in a timely manner before requesting an Admission Decision under Section 9.

(A) Examiners will review components of a Quarterly Portfolio that is submitted late, and will score components of the Portfolio, but they will score the Quarterly Portfolio itself "unqualified" due to lateness. Provisional Licensees who submit a late Quarterly Portfolio, therefore, may need to devote additional time to the Program.

Commented [JK24]: So they would re-do that (whole) quarter? Or, just resubmit (the same work) in advance of the next due date?

(B) A Provisional Licensee may request an extension of time for filing a Quarterly Portfolio by completing the "Extension of Time" form and emailing it to admissions@osbar.org at least 24 hours before the Quarterly Portfolio is due. Regulatory Counsel will grant requests for an extension only if unforeseen circumstances prevent the Provisional Licensee's timely submission. If an emergency occurs during the final 24 hours before a Quarterly Portfolio is due, the Provisional Licensee should submit the "Extension of Time" form and explain the sudden nature of the unforeseen circumstances.

7.4 Review and Grading of Portfolios. Each Quarterly Portfolio will be reviewed and graded by an Examiner. Components of the Portfolio will be scored in the following manner:

Commented [JK25]: If possible, might be better to have more than one examiner review/grade to be able to measure consistency/agreement.

(A) The "Learning the Ropes" CLE program will be scored "in progress" until the Provisional Licensee has finished the program. When the Provisional Licensee has completed the program, this component will be scored "qualified." I.e., the Examiner does not make any independent assessment of the Provisional Licensee's work in this program.

(B) The New Lawyer Mentoring Program will be scored "in progress" until the Provisional Licensee has finished the activities required by Rule 6.4. When the Provisional Licensee has completed those activities, this component will be scored "qualified." I.e., the Examiner does not make any independent assessment of the Provisional Licensee's work in this program.

Commented [JK26]: Will C--E be compensatory or non-compensatory within a category and/or across categories? For example, could the Written Work Product be judged "qualified" as a whole even if on or more of the individual documents were scored "not qualified?" Do you have to be (fully) qualified across all 12? Not understanding how the scoring works yet.

(C) Written work product will be scored "qualified" or "not qualified" using Program rubrics. The Examiner will score these documents based on independent review of the

Commented [JK27]: Just wondering how it will be "independent" if the Supervisor's rubric is included? Better to not include the supervisor's rubric if you want the judgment to truly be independent?

document, review of the Supervising Attorney's completed rubric, and review of the Provisional Licensee's accompanying reflection.

(D) Client interviews or client counseling sessions will be scored "qualified" or "not qualified" using Program rubrics. The Examiner will score this work based on the Supervising Attorney's completed rubric and review of the Provisional Licensee's reflection.

(E) Negotiations will be scored "qualified" or "not qualified." The Examiner will score this work based on the Supervising Attorney's completed rubric and review of the Provisional Licensee's reflection.

(F) Reflections will be scored "qualified" as long as they are submitted and complete. The reflections are required but their primary purpose is to help the Examiner assess the above items. The Examiner does not need to score the contents of the reflections.

(G) The learning plan will be scored "qualified" or "not qualified" using Program rubrics.

(H) Timesheets will be scored "qualified" as long as they appear reasonably complete.

(I) The Quarterly Portfolio itself will be scored "qualified" or "not qualified" according to Program rubrics.

7.5 Anonymous Grading. All Portfolios will be graded anonymously. Provisional Licensees will identify their Portfolios with numbers assigned by the Admissions Department.

7.6 Feedback. Provisional Licensees will receive the Examiner's completed rubrics, as well as a Quarterly Summary Sheet, after each Quarterly Review. The Board will make every effort to provide this feedback within 3 calendar weeks of the Quarterly Portfolio submission.

7.7 Challenges to Quarterly Assessments. Provisional Licensees may not challenge or appeal any assessments made during the Quarterly Reviews. These assessments will contribute to the final Licensing Score, but they are revealed primarily to help Provisional Licensees improve their performance.

(A) If a Provisional Licensee disagrees with an assessment, they should discuss the disagreement with their Supervising Attorney, another attorney working for the Employer, their Mentor, or an Ombudsperson. Those individuals may be able to help the Provisional Licensee understand the assessment and work to improve their performance.

(B) If a Provisional Licensee continues to disagree with an assessment, they may include a note explaining their disagreement in the Portfolio. Examiners will read that note during subsequent reviews and may choose to amend an assessment.

(C) Similarly, a Supervising Attorney who disagrees with an assessment may add a note to the Provisional Licensee's Portfolio explaining that disagreement. These notes are particularly helpful if the Supervising Attorney clarifies an aspect of the underlying matter that may not have been apparent to the Examiner reviewing the Quarterly Portfolio.

Commented [JK28]: (As noted several times throughout this section, don't understand how the reflection will be helpful to the Examiner and seems it could be biasing.)

Commented [JK29]: (Don't understand how this helps.)

Commented [JK30]: Just noting that this sounds worryingly underspecified for a licensure decision in most contexts... can any additional precision/definition be added? With so much "latitude" it might be challenging to achieve interrater reliability and/or be hard to later defend decisions.

Commented [JK31]: If not qualified, do they repeat that whole quarter? Revise and resubmit? Not sure I understand the process very fully yet.

Commented [JK32]: This goes back to my question about what exactly the role/task of the Examiner is. Will the identity of both the Candidate and the Supervisor be hidden from the Examiner? Is the Examiner mostly "agreeing or disagreeing" with whatever the Supervisor already decided, or are they rendering a fully independent judgment according to different criteria? Is it the Examiners job to do some sort of linkage exercise back to a Practice Analysis, or will this be done separately? Is the Examiner functionally setting the standard themselves (by finding the qualified/not line?) or will there be some criterion-referenced way of deciding? If the judgment is to be fully independent, the Examiner probably should not see the reflections and/or the Supervisor's evaluation(s) as part of the interim and/or final portfolio. (Also recommend considering only reviewing Final Portfolio.)

Commented [JK33]: ? Confused about several parts of this. Are we saying they can't challenge/appeal at the quarterly assessment, but can challenge/appeal the same score if/when it is part of the final portfolio? Also confused as to whether/how other attorneys would even know what scores had been assigned by the Examiners, let alone understand how to talk to the candidates about them?

Commented [JK34]: We've said in the first sentence of 7.7 that P.L's can't challenge, but this sounds like a challenge(?) I think it's important to be very clear/transparent with students about whether or not they can (actually) challenge, what that looks like, and what the possible outcomes are (amendment of previous assessment? Possible amendment?).

Commented [JK35]: Not fully clear to me why/when the Supervisor and/or Candidate would need to see the individual evaluations made by the Examiner.

7.8 Replacement of Portfolio Components. Provisional Licensees may replace some components of the Portfolio if that component has been marked "not qualified." These provisions govern those replacements:

(A) Provisional Licensees may replace up to 4 pieces of the independently authored written work product described in Rule 6.5. In other words, Provisional Licensees may submit up to 12 pieces of written work product to satisfy the requirement of 8 documents marked "qualified."

(B) Provisional Licensees may replace the materials relating to one or both client interviews or counseling sessions described in Rule 6.6. In other words, Provisional Licensees may submit materials relating to up to 4 client interviews or counseling sessions to satisfy the requirement of 2 sets of materials marked "qualified."

(C) Provisional Licensees may replace the materials relating to one or both negotiations described in Rule 6.7. In other words, Provisional Licensees may submit materials relating to up to 4 negotiations to satisfy the requirement of 2 sets of materials marked "qualified."

(D) If a Provisional Licensee replaces any of the materials discussed in subsections (A) through (C) above, they should also replace the reflections related to those materials.

(E) The replacements described in subsections (A) through (D) above may be included in any Quarterly or Final Portfolio submitted after the original component was marked "not qualified."

(F) When the Provisional Licensee adds the replacement component to the Portfolio, they may remove the original component marked "not qualified." The Final Admission Decision described in Section 9 below will not include the removed component.

(G) Learning plans, timesheets, and Quarterly Portfolios may not be replaced. Instead, as provided in Rule 9.4, Provisional Licensees must receive a "qualified" mark on just 3 learning plans, 3 timesheets, and 3 Portfolios (including the Final Portfolio and its materials) to qualify for admission to the Oregon State Bar. This allows Provisional Licensees to replace deficient learning plans, timesheets, and Quarterly Portfolios by remaining in the Program for an additional quarter.

Commented [JK36]: (most)

Commented [JK37]: If it was marked "not qualified" as part of the Final Portfolio, would there be a second (or third) "Final Portfolio?"

Commented [JK38]: If I fail to get enough(?) (all?) "qualified" ratings even after an additional quarter, what happens? Do I need to restart fully with a new supervisor? Can I "start over" with the same supervisor? Is this path closed to me (permanently/indefinitely)?

Section 8
Accommodations

8.1 Accommodations. This Program has been designed using the principles of universal design, so it should be accessible to all individuals who qualify as Provisional Licensees. If, however, a Provisional Licensee has a disability as defined under the Americans with Disabilities Act of 1990 (42 USC § 12101 et seq.), and the Provisional Licensee believes that the disability will impair their ability to complete any Program requirements, the Licensee may request reasonable accommodations. The Provisional Licensee should use the process outlined in RFA 5.10 to seek those accommodations. The filing deadline for any request filed under this provision is the date upon which the Provisional Licensee files their application to participate in

Comments from Danette McKinley for SPP Chapters 6 - 8

Section 6
Program Requirements

6.1 Overview of Program Requirements. The Program has ten substantive requirements and an hours requirement, each described in more detail in the Rules below. The substantive requirements are:

(A) Diligent, competent, and professional work on all Legal Work assigned to the Provisional Licensee by their Supervising Attorney;

(B) Completion of the Professional Liability Fund's "Learning the Ropes" CLE program;

(C) Completion of the New Lawyer Mentoring Program in the manner described in Rule 6.4;

(D) Production of at least 8 independently authored pieces of written work product;

(E) Leadership of at least 2 initial client interviews or client counseling sessions;

(F) Leadership of at least 2 negotiations;

(G) Reflections on each of the 8 pieces of written work product, 2 client interviews or counseling sessions, and 2 negotiations;

(H) Development and maintenance of a learning plan for accomplishing the above activities;

(I) Completion of regular timesheets recording all time devoted to the Program; and

(J) A Portfolio organizing the above Program components.

Commented [DM1]: Will maximums be set or will examiners review all of these materials?

Commented [DM2]: This is part of the provisional program. Needed for the SPP?

6.2 Legal Work. Provisional Licensees will perform Legal Work assigned to them by their Supervising Attorney.

(A) Legal Work, as defined by Rule 1.2(I), includes any work that is commonly performed by licensed attorneys in Oregon. Legal work may include activities that are also performed by unlicensed individuals, as long as newly licensed attorneys regularly incorporate those activities in their work.

(B) Provisional Licensees must perform this work diligently, competently, and professionally.

(C) A Provisional Licensee should not attempt work for which they feel unprepared or incompetent to perform. Instead, the Provisional Licensee should discuss their reservations with the Supervising Attorney and seek appropriate assistance.

(D) Provisional Licensees may request particular types of Legal Work that would benefit their professional development or completion of this Program. Provisional Licensees may also request permission to work on pro bono matters. If a Supervising Attorney agrees to any of these requests, that agreement constitutes "assignment" of Legal Work.

6.3 Learning the Ropes. The Provisional Licensee must attend or watch all sessions of the 2021 or 2022 "Learning the Ropes" program offered by the Oregon State Bar Professional Liability

Fund. When the Provisional Licensee has satisfied this requirement, they must include their certificate of completion in their Portfolio.

6.4 New Lawyer Mentoring Program (NLMP). The NLMP Manual explains all required and optional elements of the NLMP.

(A) Provisional Licensees must follow the directions in the Manual for completing the required activities in Parts 1 to 5; at least one optional activity in each of those Parts; and at least 10 practice activities.

(B) Provisional Licensees must use the Manual worksheets to document their completion of required activities, optional activities, and practice activities. Those worksheets must be included in each Quarterly Portfolio, updated to reflect all activities completed to date.

(C) Provisional Licensees may not use any of the required activities specified in Rules 6.5 to 6.11 to satisfy the required activities, optional activities, or practice activities required by the NLMP. All NLMP activities are in addition to the other requirements of this Program.

(D) Lawyers who obtain their permanent licenses through the Provisional Licensing Program will not have to repeat the NLMP during their first year of practice. That requirement will be waived for lawyers completing the NLMP as part of the Provisional Licensing Program.

6.5 Independently Authored Written Work Product.

(A) Each Provisional Licensee must prepare and submit at least 8 pieces of independently authored work product. Written work product may take any form that lawyers use in their practices: memos, letters, emails, complaints, motions, briefs, contracts, wills, etc. All submitted work product, however, must comply with the following requirements:

(1) The work product must address some *substantive* aspect of a legal matter. Provisional Licensees may not submit routine scheduling notes or engagement letters for this component of the Portfolio.
(2) At least 2 of the pieces of work product must exceed 1500 words, not including headers or signature blocks. Footnotes do count towards the word total.
(3) All submitted work product must be *independently* authored by the Provisional Licensee. If another individual edits the Provisional Licensee's work, the Provisional Licensee must submit the original draft without those edits.

If Provisional Licensees are unsure whether a particular piece of work product qualifies for submission, they may ask one or both Program Managers. The Program Managers will not judge the quality of the work product, but can advise the Provisional Licensee whether the work product meets these minimum requirements.

(B) Each piece of work product must be accompanied by a brief assessment completed by the Supervising Attorney. The Board will provide standardized rubrics to use for these assessments.

Commented [DM3]: How will the Board use the assessment results?

(C) Each piece of work product must also be accompanied by the Provisional Licensee's structured reflection. The Board will provide a standardized template for that reflection.

Commented [DM4]: What is the purpose of sharing the reflection? Will that contribute towards a licensure decision?

(D) If the Provisional Licensee used a sample, template, or other work as foundation for the work product, the Provisional Licensee must also:

(1) Submit a copy of the original sample, template, or other work used as a foundation; and
(2) Highlight the portions of the work product that represent the Provisional Licensee's additions, edits, or other customization.

(E) If the work product relates to a client matter, and the work product has not been publicly filed in substantially the same form:

(1) The work product must be redacted to protect the client's interests; and
(2) The client must consent to inclusion of the work product in the Portfolio.

(F) If the Provisional Licensee is unable to gather sufficient work product from client-related work, the Supervising Attorney may assign a mock exercise or exercises to the Provisional Licensee, which can be submitted to fulfill this requirement. All mock work product must comply with subsections (A) through (D) above.

6.6 Client Interviews or Counseling Sessions. Each Provisional Licensee must lead two client interviews or counseling sessions that are observed and assessed by their Supervising Attorney.

(A) Requirements for the Provisional Licensee:

(1) Before beginning each interview or counseling session, the Provisional Licensee must explain the Supervising Attorney's role, assure the client that the Supervising Attorney (a member of the Provisional Licensee's organization) is encompassed by the attorney-client privilege, and obtain the client's consent to the Supervising Attorney's presence.
(2) If the client objects to the Supervising Attorney's presence, the Supervising Attorney will depart, and the Provisional Licensee will conduct the session without assessment. The Provisional Licensee must find another interview or counseling session to satisfy this component of the Program.
(3) After completing the interview or counseling session and receiving feedback from the Supervising Attorney, the Provisional Licensee must complete a structured reflection about the interview or counseling session. The Board will provide a template for that structured reflection.

(B) Role of the Supervising Attorney:

(1) The Supervising Attorney should not speak during the interviews or counseling sessions other than to greet the client at the beginning of each session, provide input specifically requested by the Provisional Licensee, or intervene if necessary to protect the interests of the client.
(2) If the client addresses a question to the Supervising Attorney, the Supervising Attorney should politely redirect the question to the Provisional Attorney.

(3) For each of the two interviews or counseling sessions, the Supervising Attorney will complete the client interview/counseling session rubric provided by the Board, share the completed rubric with the Provisional Licensee, and offer any additional feedback that would assist the Provisional Licensee's development.
(4) The Supervising Attorney's completed rubric will become part of the Provisional Licensee's Portfolio.

(C) Exceptions:

(1) Provisional Licensees working in a prosecutor's office may use interviews or other sessions with a complainant to satisfy this component of the Program. The provisions in subsections (A) and (B) will govern those sessions except that the attorney-client privilege will not apply.
(2) If a Provisional Licensee does not conduct client (or complainant) interviews or counseling sessions in their workplace, the Program Managers will help the Provisional Licensee find other ways to satisfy this component of the Program. It may be possible, for example, for the Provisional Licensee to interview simulated clients as part of a law school competition or CLE exercise. Any substitutes will be structured to follow the other portions of this Rule as closely as possible.

6.7 Negotiations. Each Provisional Licensee must conduct two negotiations that are observed and assessed by the Supervising Attorney. A negotiation includes any discussion aimed at reaching an agreement. It may occur in the context of litigation, transactional, regulatory, or other matters. The negotiation does not have to focus on final resolution of the matter; it may focus on preliminary or interim matters.

(A) Requirements for the Provisional Licensee:

(1) Before beginning each negotiation, the Provisional Licensee must explain the Supervising Attorney's role to other attorneys participating in the negotiation. If the client is present, the Provisional Licensee must also explain the Supervising Attorney's role to the client and obtain the client's consent to the Supervising Attorney's presence.
(2) If the client objects to the Supervising Attorney's presence, the Supervising Attorney will depart, and the Provisional Licensee will conduct the negotiation without that assessment. The Provisional Licensee must find another negotiation to satisfy this component of the Program.
(3) After completing the negotiation and receiving feedback from the Supervising Attorney, the Provisional Licensee will complete a structured reflection about the negotiation. The Board will provide a template for that structured reflection.

(B) Role of the Supervising Attorney:

(1) The Supervising Attorney should not speak during the negotiations other than to greet other attorneys and the client at the beginning of each session, provide input specifically requested by the Provisional Licensee, or intervene if necessary to protect the interests of the client.

(2) If another attorney or the client addresses a question to the Supervising Attorney, the Supervising Attorney should politely redirect the question to the Provisional Attorney.

(3) For each of the two negotiations, the Supervising Attorney will complete the negotiation rubric provided by the Board, share the completed rubric with the Provisional Licensee, and offer any additional feedback that would assist the Provisional Licensee's development. The completed rubric will become part of the Provisional Licensee's Portfolio.

(C) Exceptions: If a Provisional Licensee does not conduct negotiations in their workplace, the Program Managers will help the Provisional Licensee find other ways to satisfy this component of the Program. It may be possible, for example, for the Provisional Licensee to conduct negotiations as part of a law school competition or CLE exercise. Any substitutes will be structured to follow the other portions of this Rule as closely as possible.

6.8 Reflections. The Provisional Licensee must complete reflections, using templates provided by the Board, for (a) each of the 8 independently authored written work products submitted under Rule 6.5; (b) each of the 2 client interviews or counseling sessions conducted under Rule 6.6; and (c) each of the 2 negotiations conducted under Rule 6.7. The Provisional Licensee must include all of these reflections in their Portfolio.

6.9 Learning Plan. The Learning Plan has two purposes: (a) It tracks what the Provisional Licensee has accomplished to date, and (b) it sets goals for the coming quarter.

(A) The Board will provide a Learning Plan template to help Provisional Licensees create and update their Learning Plans.

(B) Provisional Licensees must update the Learning Plan regularly and include the current Plan in each Quarterly Portfolio.

(C) The final Learning Plan, submitted as part of the Final Portfolio, will demonstrate how the Provisional Licensee has accomplished all requirements of the Program. It will also require the Provisional Licensee to reflect on how they will continue to improve their skills and knowledge as a practicing lawyer.

6.10 Timesheets. Provisional Licensees must complete timesheets for every day spent working in the Program. These timesheets are separate from any timekeeping required by the Supervising Attorney or Employer. The Program timesheets are used to document achievement of the 1500 hours required by the Program and to demonstrate the Provisional Licensee's competence in timekeeping.

(A) The Board will provide timesheet blanks for Provisional Licensees to use for these purposes.

(B) Provisional Licensees may use those timesheets to record their time in either 6- or 15-minute increments, whichever is most convenient for them.

(C) Decimals must be used to record all time. E.g., 30 minutes is .5 hours.

(D) Provisional Licensees should record all time devoted to the Program, even if that time is not billable to a client.

(E) Timesheets must be submitted as part of both Quarterly Portfolios and the Final Portfolio.

6.11 Portfolio. The Provisional Licensee must create and maintain a Portfolio collecting all of the above materials. The Board will provide a template for organizing the Portfolio. As explained in Section 7 below, Provisional Licensees must submit their Portfolio for review at the end of every quarter they participate in the Program. When the Provisional Licensee is ready to seek admission to the Oregon State Bar, the Provisional Licensee must prepare and submit a Final Portfolio as provided in Rule 9.2.

6.12 Hours. To qualify for admission to the Oregon State Bar, Provisional Licensees must document at least 1500 hours spent working within the Provisional Licensing Program. Those hours may include:

(A) All time spent devoted to Legal Work assigned by the Supervising Attorney, even if the time is not billed to a client;

(B) All time devoted to working on the Program components outlined in Rules 6.3 to 6.11.

(C) All time spent discussing or learning about Program components with their Supervising Attorney, other Employees of the Employer, their Mentor, the Program Managers, or an Ombudsperson;

(D) All time spent reviewing or reflecting on feedback from the Board on Quarterly Portfolios;

(E) All time spent in any training or educational activities required by their Employer that are not included in the Program components; and

(F) Up to 30 additional hours of MCLE activities.

Section 7
Quarterly Portfolios and Review

7.1 Requirement of Quarterly Portfolios. Provisional Licensees must submit their Portfolios to the Board each quarter that they participate in the Program. These Quarterly Portfolios must contain all Program work described by Rules 6.3 to 6.10 that has been completed to date.

Commented [DM5]: Is the purpose to monitor progress towards the minimal documents in Section 6? Why are examiners 'scoring' – they will make qualified-not qualified decisions.

7.2 Notification of Quarterly Due Dates. Regulatory Counsel will notify Provisional Licensees of the time and date when each Quarterly Portfolio is due, as well as the process for submitting the Quarterly Portfolio. That notice will be provided at least 30 calendar days before the due date.

7.3 Timeliness of Quarterly Portfolios. Lawyers frequently must comply with deadlines, and missing a deadline can have serious consequences for a client. Provisional Licensees, therefore, must submit at least 2 Quarterly Portfolios in a timely manner before requesting an Admission Decision under Section 9.

(A) Examiners will review components of a Quarterly Portfolio that is submitted late, and will score components of the Portfolio, but they will score the Quarterly Portfolio itself "unqualified" due to lateness. Provisional Licensees who submit a late Quarterly Portfolio, therefore, may need to devote additional time to the Program.

Commented [DM6]: This clarifies that the decision is not pass-fail – just whether or not the materials qualify. Is there a way to distinguish this in the rules?

(B) A Provisional Licensee may request an extension of time for filing a Quarterly Portfolio by completing the "Extension of Time" form and emailing it to admissions@osbar.org at least 24 hours before the Quarterly Portfolio is due. Regulatory Counsel will grant requests for an extension only if unforeseen circumstances prevent the Provisional Licensee's timely submission. If an emergency occurs during the final 24 hours before a Quarterly Portfolio is due, the Provisional Licensee should submit the "Extension of Time" form and explain the sudden nature of the unforeseen circumstances.

7.4 Review and Grading of Portfolios. Each Quarterly Portfolio will be reviewed and graded by an Examiner. Components of the Portfolio will be scored in the following manner:

(A) The "Learning the Ropes" CLE program will be scored "in progress" until the Provisional Licensee has finished the program. When the Provisional Licensee has completed the program, this component will be scored "qualified." I.e., the Examiner does not make any independent assessment of the Provisional Licensee's work in this program.

(B) The New Lawyer Mentoring Program will be scored "in progress" until the Provisional Licensee has finished the activities required by Rule 6.4. When the Provisional Licensee has completed those activities, this component will be scored "qualified." I.e., the Examiner does not make any independent assessment of the Provisional Licensee's work in this program.

(C) Written work product will be scored "qualified" or "not qualified" using Program rubrics. The Examiner will score these documents based on independent review of the document, review of the Supervising Attorney's completed rubric, and review of the Provisional Licensee's accompanying reflection.

(D) Client interviews or client counseling sessions will be scored "qualified" or "not qualified" using Program rubrics. The Examiner will score this work based on the Supervising Attorney's completed rubric and review of the Provisional Licensee's reflection.

(E) Negotiations will be scored "qualified" or "not qualified." The Examiner will score this work based on the Supervising Attorney's completed rubric and review of the Provisional Licensee's reflection.

(F) Reflections will be scored "qualified" as long as they are submitted and complete. The reflections are required but their primary purpose is to help the Examiner assess the above items. The Examiner does not need to score the contents of the reflections.

(G) The learning plan will be scored "qualified" or "not qualified" using Program rubrics.

(H) Timesheets will be scored "qualified" as long as they appear reasonably complete.

(I) The Quarterly Portfolio itself will be scored "qualified" or "not qualified" according to Program rubrics.

Commented [DM7]: How will the quarterly decisions contribute to an overall decision?

7.5 Anonymous Grading. All Portfolios will be graded anonymously. Provisional Licensees will identify their Portfolios with numbers assigned by the Admissions Department.

7.6 Feedback. Provisional Licensees will receive the Examiner's completed rubrics, as well as a Quarterly Summary Sheet, after each Quarterly Review. The Board will make every effort to provide this feedback within 3 calendar weeks of the Quarterly Portfolio submission.

7.7 Challenges to Quarterly Assessments. Provisional Licensees may not challenge or appeal any assessments made during the Quarterly Reviews. These assessments will contribute to the final Licensing Score, but they are revealed primarily to help Provisional Licensees improve their performance.

(A) If a Provisional Licensee disagrees with an assessment, they should discuss the disagreement with their Supervising Attorney, another attorney working for the Employer, their Mentor, or an Ombudsperson. Those individuals may be able to help the Provisional Licensee understand the assessment and work to improve their performance.

(B) If a Provisional Licensee continues to disagree with an assessment, they may include a note explaining their disagreement in the Portfolio. Examiners will read that note during subsequent reviews and may choose to amend an assessment.

(C) Similarly, a Supervising Attorney who disagrees with an assessment may add a note to the Provisional Licensee's Portfolio explaining that disagreement. These notes are particularly helpful if the Supervising Attorney clarifies an aspect of the underlying matter that may not have been apparent to the Examiner reviewing the Quarterly Portfolio.

7.8 Replacement of Portfolio Components. Provisional Licensees may replace some components of the Portfolio if that component has been marked "not qualified." These provisions govern those replacements:

(A) Provisional Licensees may replace up to 4 pieces of the independently authored written work product described in Rule 6.5. In other words, Provisional Licensees may submit up

to 12 pieces of written work product to satisfy the requirement of 8 documents marked "qualified."

(B) Provisional Licensees may replace the materials relating to one or both client interviews or counseling sessions described in Rule 6.6. In other words, Provisional Licensees may submit materials relating to up to 4 client interviews or counseling sessions to satisfy the requirement of 2 sets of materials marked "qualified."

(C) Provisional Licensees may replace the materials relating to one or both negotiations described in Rule 6.7. In other words, Provisional Licensees may submit materials relating to up to 4 negotiations to satisfy the requirement of 2 sets of materials marked "qualified."

(D) If a Provisional Licensee replaces any of the materials discussed in subsections (A) through (C) above, they should also replace the reflections related to those materials.

(E) The replacements described in subsections (A) through (D) above may be included in any Quarterly or Final Portfolio submitted after the original component was marked "not qualified."

(F) When the Provisional Licensee adds the replacement component to the Portfolio, they may remove the original component marked "not qualified." The Final Admission Decision described in Section 9 below will not include the removed component.

(G) Learning plans, timesheets, and Quarterly Portfolios may not be replaced. Instead, as provided in Rule 9.4, Provisional Licensees must receive a "qualified" mark on just 3 learning plans, 3 timesheets, and 3 Portfolios (including the Final Portfolio and its materials) to qualify for admission to the Oregon State Bar. This allows Provisional Licensees to replace deficient learning plans, timesheets, and Quarterly Portfolios by remaining in the Program for an additional quarter.

Section 8
Accommodations

8.1 Accommodations. This Program has been designed using the principles of universal design, so it should be accessible to all individuals who qualify as Provisional Licensees. If, however, a Provisional Licensee has a disability as defined under the Americans with Disabilities Act of 1990 (42 USC § 12101 et seq.), and the Provisional Licensee believes that the disability will impair their ability to complete any Program requirements, the Licensee may request reasonable accommodations. The Provisional Licensee should use the process outlined in RFA 5.10 to seek those accommodations. The filing deadline for any request filed under this provision is the date upon which the Provisional Licensee files their application to participate in the Program, although Provisional Licensees may file later requests if the disability occurs or the need for accommodation becomes apparent after the filing deadline.

EXHIBIT 2

Comments from Joanne Kane for OEP

OEP Admissions Qualifications/Selection

From the original Supreme Court submission:

> Finally, we recommend expressly encouraging holistic admission practices[1] including admitting law students[2] on more than an evaluation of LSAT/GPA[3] in order to ensure reliance on more inclusive criteria, such as work experience, life experience, and/or overcoming personal challenges[4]. Law schools will inherently be encouraged to do so[5] if they have the confidence that all first-year students can apply for the OEP program[6]. Accordingly, we recommend making clear that the OEP will be open to all students in the spring[7] of 1L year[8] (rather than limiting participation to those pre-selected for the program)

---

[1] Might be helpful to spell out what this means, and/or what it doesn't mean. Are we using "holistic" to mean "considering more than just law school GPA earned during 1L, but still somewhat standardized?" Or are we using "holistic" to mean there will not (by design) be standardization across schools, time, decision-makers, etc.? Some scholars have argued that holistic admission does more harm than good (see: https://www.theatlantic.com/education/archive/2013/12/the-false-promise-of-holistic-college-admissions/282432/). Is there any reassurance we can offer here of doing more good than harm? Specifically to avoid the concern that, "it's a way for [a program] to discreetly take various sensitive factors..., [e.g.,] students whose families might donate a gym – into account." Is there a middle ground where we can be transparent about what is being considered and have many things be considered in addition to or in lieu of tests/grades? (E.g., work experiences, whether those are relevant to the practice of law; volunteer experiences, whether those are relevant to the practice of law; fluency in multiple languages including English; ability to participate in and/or pay for other pathways; etc. etc.?) Not intending to cast doubt on any programs/potential decision makers, but it seems like at least some transparency would go a long way toward both perceived and actual fairness.

[2] It sounded at the meeting like we needed to clarify that we meant admitting to the OEP program, not to law school.

[3] Are we thinking this would be LGPA after first year? Or after just first semester?

[4] Just a reminder that we noted at the meeting we need to clarify whether these are initial criteria (when far more students want to participate than can participate given limited spots), or whether these are criteria designed to be used even after the program is at scale.

[5] Just noting I'm not sure what this sentence means. Can we try to rewrite a little to help me/other readers better understand this part? I know this was taken from a previous report, but understanding it might help us get on the same page and/or help clarify some points in the next report.

[6] It might be good to clarify whether the ultimate goal with this program is that every law student in OR would (or at least could, if they wanted to) participate in this, or whether the numbers will be kept lower more similar to the Daniel Webster Scholars program. It had sounded in some previous conversations like the goal was to develop a program in which every law student in OR could participate, at least after a ramp-up period.

[7] Mention implications for 1L curriculum? (If any?)

[8] There seemed to be good consensus around this when we briefly discussed, but subsequently I've wondered if it'd be fairer to let people know whether they are accepted into the program before they start LS. I'm concerned about applicants who hear about this pathway, come to law school expecting to get accepted into the pathway, but then actually aren't accepted and have to follow another pathway -- even though they are unlikely to be able to find a supervisor (few connections, etc.) or to pass the bar exam (low law school GPA, which is the best predictor of bar exam passage.) If the criteria used focus on work experience, life experience, and/or overcoming personal challenges as suggested in the original Supreme Court submission, (most of) that could likely be known

- Do not have general statements about law school admissions
    - Requiring or prescribing admissions processes
    - How might OEP change their regular admissions process

Common Qualifications:

- GPA: Good academic standing vs. number (not uniform)[9]
- Writing skills[10]
- Process for holistic review of candidates
- Foundational skill qualifications (lawyer vs law student) (possibly evidence based)
- Consider difference between a helpful list vs subjective indicators

Requirements[11]:

---

for most applicants before law school. Perhaps a few more opportunities could open after 1L for students who demonstrated need/interest/fit, but most spots would be filled before 1L so students knew what they were signing up for before they invested time and money and built up other "sunk costs" their first year.

[9] I wonder, rather than basing on GPA at all, whether this should simply be a lottery, at least for the first few years. That would take away the concerns in the first footnote, about whether people were being selected based on "unfair" factors (like donating a gym) other than dedication to the profession and/or unique ability to flourish within the program, or whatever factors we would like to consider and like to say had been taken into consideration. If it's not done as a lottery, would we want to say people with the *lowest* GPA would be most likely selected for this program, given the LPDC's goal of targeting people who are simply bad test takers and receive low law school grades but would still be very good lawyers? If that is the group of people the LPDC has had in mind, seems like the criteria developed should be designed to favor that group. The Daniel Webster Scholars curricular pathway that includes "specific, intensive" courses based around "simulated, clinical, and legal residency settings" – such individualized, hands-on learning opportunities would likely benefit most students, but maybe especially students who do need more individualized support. (https://law.unh.edu/academics/daniel-webster-scholar-honors-program#:~:text=GPA%3A%20Webster%20Scholars%20must%20graduate,any%20DWS%2Donly%20designated%20course.) *Important note: if students are selected based on low GPA, efforts might need to be made to prevent this becoming public knowledge (since it could be stigmatizing), and, it's a little strange to imagine, but maybe steps would need to be taken to prevent a sort of a "sandbagging" approach where students would strategically do poorly in their 1L year to be eligible for this if there are limited opportunities to participate? This would require careful thought and careful communications for sure.*

[10] Will this have the unintended consequence of making the 1L writing classes far more competitive and high-stakes? Or, perceived by students as high-stakes and more pressured/stressful? Is this a concern?

[11] A key question here is the extent to which, in both the short and long terms, resources will be available to effectively "staff" the roles need to fulfill the instructional demands? I know the LPDC felt that the cost of the current bar exam ($900 or less for a first-time taker – and bar prep materials - $125 for NCBE's bar study materials) was considered by many members to be prohibitively expensive. Even if you calculate the costs of the bar exam differently (by adding on multiple bar review courses and/or opportunity costs, for example), this seems like it still will almost certainly be more expensive unless there is a large volunteer component or similar. It'd be helpful to see costs laid out and compared across all the pathways, along with information about who will bear the costs, given reducing the financial burden on students is a clear (and worthy!) goal of the initiative. Will the additional costs be borne by the students participating in the program? By all students? Through another funding source?

Documentation[12]:

Advising:

Entry: Preliminary program plan

Ongoing: Regular check-ins to measure ongoing success/challenges in program

Meeting requirements to continue program

Outboarding

Others:

[12] When thinking about the documentation/factors tracked in association with this program, it'd be helpful to go back to the goals/impetus for the program. We've heard from the LPDC members that one particularly motivating factor for this work is that there are differences by demographic group in LSAT scores, undergraduate GPA, and law school GPA on average - and since all of those predict bar exam scores, you see the same trends on the exam outcomes. In recognition of that (again, excellent) goal of eliminating differences in participation/passage by group, it'd be important to show that there were smaller, or ideally **no** differences by group in who was being selected for the program and who was completing the program. *Or* to show that people from historically underrepresented groups were *more* likely to be selected and—crucially—equally or more likely to complete the process/program and become licensed. This group should probably think about who will be responsible for tracking/reporting these outcomes. (Schools? BBX? Both? Other?)

Comments from Danette McKinley for OEP

Troy Wood

From: Danette McKinley <dmckinley@ncbex.org>
Sent: Thursday, September 15, 2022 9:20 AM
To: Troy Wood
Subject: OR licensure pathways - OEP and SPP
Attachments: Psychometric Guidelines IAALS_NCBE.pdf; Pommerich Fairness Comparing Test Scores.pdf

Hello Troy.

I usually go through my committee but wanted to reach out to you directly with comments that I believe will apply to both the curricular and supervised practice pathways. I wanted to share my thoughts about the OEP. I have not attended those meetings, but have been discussing their progress with Joanne Kane, Logan Cornett and Deb Merritt. Now that Addie, Tony and Deb have met to discuss coordination of the reports the subcommittees will submit, I thought it was a good time to share my thoughts with you.

Here's what I was thinking about the educational competence pathway.

1. The schools would work closely to show how the curricular objectives meet the requirements set by the regulators (the bar). I suspect that there are aspects of this already in place.
2. In this environment, there is the opportunity for the schools to develop assessments that directly address competencies defined as necessary for general, unsupervised practice upon graduation. The decision is not whether the students are ready or not (that's up to the examiners). Instead, these assessments indicate progression towards that goal. In this way it would be similar to entrusted professional activities in medicine – where entrustment develops over the course of the educational program. These assessments would not be part of the portfolio – I'm thinking they could be years 1 and 2.
3. Year 3 is where the materials to submit could be generated. There could be two examinations consisting of simulations that represent the various types of practice areas and that would focus on skills measurement. Whether these simulations produce transcripts or not is something to be decided. I think that this would be similar to my concerns about the supervised practice competence pathway – that we allow the examiners to make bar admissions decisions. What the faculty is responsible for is making different judgments about whether the student is learning and progressing through the program towards graduation.
4. The writing component seems as if it will be challenging – judgment needs to be targeted at the right point. The extent to which the submissions to any type of portfolio are independent is essential. What happens to the lawyer in solo practice? Does that attorney draft a document and then revise it themselves? Are documents regularly shared and revised?
5. Jumping ahead to the assessments, I think that simulations can be well designed for assessment – and the examiners may know the students. That is actually how objective structured clinical exams have been administered at medical schools for years. The examiners are faculty in the program and they rate the students on various aspects of how they manage the patients in the encounters. Patients (standardized) are also asked to provide ratings of communication skills. In fact, the medical school OSCEs are often given periodically for formative feedback, with a different set administered for the high stakes end-of-year exams. This could mean that there are simulations graded by faculty. That does not mean that the graded materials are sent to BBX – those examiners are making a different decision.

More generally, it is important to plan to accumulate evidence for the validity of the results of the various pathways used for bar admissions. I was encouraged by Deb Merritt's update saying that there was general agreement that there would be consistency between the curricular and supervised practice pathways. The use of similar rubrics will have a number of advantages including development time and training time for supervisors and faculty could be streamlined.

In addition to using performance samples to set the passing score (standard setting) and ensuring the definition of the minimally competent candidate is the same across all pathways, there are other details that will need to be developed in order to ensure fairness and collect validity evidence to support the admissions program. I have attached two documents. The first I believe we shared before, on psychometric criteria for assessments. The second is a chapter from Fairness in Educational Assessment and Measurement (edited by Neil Dorans and Linda Cook), by Mary Pommerich, "The Fairness of Comparing Test Scores across Different Tests or Modes of Administration".

I hope this will help us in guiding discussions on the implementation of these pathways.

Kind regards,
Danette

Danette Waller McKinley, PhD (she/her), Director of Diversity, Fairness, and Inclusion Research

National Conference of Bar Examiners
302 South Bedford Street, Madison, WI 53703-3622
215-850-1838 | dmckinley@ncbex.org | www.ncbex.org
Facebook | LinkedIn | Twitter | Instagram
Building a competent, ethical, and diverse legal profession.

Guidelines for a Licensing System Based on Supervised Practice

Logan Cornett,[1] Danette McKinley,[2] and Deborah Jones Merritt[3]

Licensing systems protect the public by certifying only candidates who are minimally competent to practice the profession. Many systems use written exams, along with other requirements, to measure that competence. Measurement experts, however, have long recognized that assessments of workplace performance can also form the core of a licensing system. "The time-honored way to find out whether a person can perform a task," one group of highly regarded experts wrote, "is to have the person try to perform the task."[4]

These experts come from the field of psychometrics, an academic field that focuses on techniques for measuring knowledge, skills, attitudes, and other facets of human cognition. Assessment experts set three primary criteria for use in the selection of licensing instruments: reliability, validity, and fairness.[5]

- Reliability means that an assessment produces consistent results. A reliable bathroom scale registers the same weight if you step on the scale twice within a minute.
- Validity means that evidence establishes a link between the assessment outcomes and the purpose for which those outcomes are used. It is valid to use bathroom scale readings to determine the body weight of able-bodied people, but not to determine what they ate for breakfast. Psychometricians stress that validity is not a property of the assessment itself, but of the interpretations made based on the scores.
- Fairness means that the assessment does not discriminate, explicitly or implicitly, based on characteristics that are irrelevant to the quality being measured. Fairness also requires assessors to treat candidates with respect. A bathroom scale is not a fair measure of body weight for a person wearing a heavy leg brace who cannot stand without that brace.

---

[1] Director of Research, IAALS (Institute for the Advancement of the American Legal System).
[2] Director of Diversity, Fairness, and Inclusion Research, National Conference of Bar Examiners (NCBE).
[3] Distinguished University Professor and John Deaver Drinko/Baker & Hostetler Chair in Law Emerita, The Ohio State University.
[4] Michael Kane, Terence Crooks & Allan Cohen, *Validating Measures of Performance*, EDUC. MEASUREMENT: ISSUES & PRACTICE, Summer 1999, at 5. Dr. Kane holds the Samuel J. Messick Chair in Test Validity at the Educational Testing Service. He previously served as Director of Research at NCBE and held faculty positions at several universities.
[5] AMERICAN EDUCATIONAL RESEARCH ASSOCIATION, AMERICAN PSYCHOLOGICAL ASSOCIATION, NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION, STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTING (2014).

Scholars have identified two other criteria that may affect the choice of licensing instruments: alignment and feasibility.[6]

- Alignment means that education and licensing align to produce new professionals with the knowledge, skills, and judgment needed for entry-level practice. For example, if negotiation is an important skill in law practice, then including negotiation tasks in a licensing assessment should motivate students to work on their negotiation skills.
- Feasibility means that the instrument used should be affordable and efficient; the test should not be too costly for either administrators or test takers. Feasibility, however, should not impede change: new approaches often seem more costly than established ones because the costs of the latter have become normalized.

A substantial body of psychometric literature shows that assessments of workplace performance can offer a reliable and valid means for determining competence. Fairness, alignment, and feasibility have been less extensively studied in the context of performance assessments, but an emerging literature suggests that these criteria can also be satisfied in that context. This handout offers some guidelines from the educational literature for designing a licensing system for lawyers based on post-graduate supervised practice. Similar principles would guide design of a system based on an experiential education path, but a future handout will address that pathway.

1. Begin with an Evidence-Based Definition of Competence. A sound assessment system rests on an evidence-based definition of the quality being measured. For a licensure system, that quality is minimum competence to practice the profession. In law, this aspect of the validity argument is based on identifying the tasks newly licensed lawyers perform, as well as the knowledge, skills, and judgment they need to perform those tasks. A national study by IAALS (Institute for the Advancement of the American Legal System), *Building a Better Bar: The Twelve Building Blocks of Minimum Competence*, identifies the competencies that newly licensed lawyers and their employers associate with entry-level practice.[7] A recent practice analysis by the National Conference of Bar Examiners (NCBE), the *Testing Task Force Phase 2 Report*, offers another helpful overview of the competencies and tasks newly licensed lawyers perform.[8] Some states have conducted similar analyses within their own jurisdiction. For example, the State Bar of California did so with a practice analysis published in 2020.[9]

---

[6] John J. Norcini & Danette W. McKinley, *Assessment Methods in Medical Education*, 23 TEACHING & TEACHER EDUC. 239 (2007).

[7] The report is available at https://iaals.du.edu/publications/building-better-bar.

[8] This report is available at https://nextgenbarexam.ncbex.org/reports/phase-2-report/.

[9] California's report is available at https://www.calbar.ca.gov/Portals/0/documents/reports/2020/California-Attorney-Practice-Analysis-Working-Group-Report.pdf.

Identifying the knowledge, skills, and judgment that new professionals use to perform common tasks is essential to define *what* needs to be measured. The definition of competence also provides guidance for *how* the requisite knowledge and skills should be measured.

2. Gather Rich Data for Each Candidate. The Uniform Bar Exam (UBE) achieves high reliability in part because it gathers many data points: answers to 175 scored multiple-choice questions, six essay responses, and two documents created for performance tests. These varied data points also support interpretations made about scores. A licensing system based on supervised practice can also achieve sufficient reliability, while supporting broader claims of competence, if it collects many data sources to support interpretations of the candidate's performance. There are several instruments that have been used in the health professions, teacher certification, and human resources field to evaluate workplace performance. This research suggests that an assessment system based on supervised practice should:

- Collect information from multiple sources (supervisors, peers, staff, the candidate, and potentially clients);
- Measure a range of relevant knowledge and skills;
- Use different formats; and
- Rely upon multiple observations.

Materials like these can be assembled into a portfolio. The richness of the data addresses reliability and allows collection of sufficient evidence of validity. Candidates can be assessed on many competencies and tasks, providing evidence to support interpretations made about competence. Multiple assessments from a wide number of raters can be monitored for potential bias, promoting fairness. Because the measures sample the full range of knowledge and skills, the multiple data sources contribute to the reliability of workplace-based assessment, supporting the decisions made about performance. With a large number of data points, errors associated with a single assessment are reduced. The portfolio as a whole is likely to paint a reliable portrait of the candidate's competence.

3. Give Candidates Feedback and Opportunities to Improve. The supervision period should provide opportunities for candidates to grow and learn from their mistakes. Supervised practice is an important part of the education process, and it is essential to provide guidance towards improvement. Initial months should focus on formative assessments designed to provide feedback rather than summative assessments used for decision-making. That focus benefits both the candidate and the employer, because the candidate will learn to provide better service.

Implementation of this guideline is particularly important in promoting fairness and reliability. Providing similar opportunities for recent graduates to hone their competence, especially for

skills and knowledge that are not taught in law school, addresses aspects of fairness in education and employment. Research, meanwhile, suggests that supervisors are more comfortable providing negative feedback if they know that candidates will have an opportunity to improve. This safety net enables assessors to give negative feedback when warranted—and to fail candidates who have not responded adequately to that feedback. Training and feedback during supervised practice also aligns with cognitive science on the development of expertise;[10] candidates are provided with opportunities to progress towards goals through practice and the provision of feedback.

4. Take Certification Out of the Workplace. Supervisors, peers, and other members of a candidate's workplace offer essential information about the candidate's competence. Indeed, an employer's willingness to allow a candidate to handle client matters offers strong evidence of the candidate's competence. It is best, however, for an independent decision maker to make the final decision about a candidate's eligibility for a license. This reinforces the acceptance of supervised practice as a learning opportunity for candidates, and supports the focus on formative feedback and the ongoing development of knowledge and skills.

This structure also lends credibility to the assessment system. Equally important, it allows candidates to change supervisors during the assessment period, which helps protect them from abusive or unethical supervisors. In these ways, the structure promotes fairness, reliability, and validity.

There are several models for independent certification. In the simplest model, a trained examiner reviews the candidate's portfolio, using established rubrics and standards, and determines whether the portfolio establishes minimum competence. If the examiner has doubts—or rules against the candidate—then a panel of examiners reviews the portfolio and reaches a consensus decision on the candidate's competence.

5. Use Credible Supervisors and Examiners. The reliability, fairness, and validity of a licensing system depend greatly on the credibility of the professionals administering the system. In a system that relies upon supervised practice, the supervisors should be licensed lawyers within the state who have demonstrated an interest in training and mentoring new lawyers. Their disciplinary records should be clean or demonstrate clear rehabilitation. A supervisor should have at least three years of experience practicing law, with at least two of those years in the state. Greater seniority does not necessarily spell better supervisory competence. Lawyers in government agencies, nonprofit organizations, and small law firms assume primary

[10] K. Anders Ericsson & Kyle W. Harwell, *Deliberate Practice and Proposed Limits on the Effects of Practice on the Acquisition of Expert Performance: Why the Original Definition Matters and Recommendations for Future Research*, FRONTIERS IN PSYCH. 10:2396 (2019).

responsibility for client matters very early in their careers. These junior lawyers may also be more in touch with new practice methods than their senior colleagues.

Examiners, similarly, should have experience practicing law in the state. It may be appropriate to require more years of experience (five or seven years) for an examiner than for a supervisor. It is best for examiners to devote only part of their time to portfolio examination, while they maintain positions practicing or teaching law. The latter work keeps them rooted in the profession, enhancing their credibility.

Diversity, equity, and inclusion are key components of credibility for both supervisors and examiners. The state's high court should assemble a group of examiners who are demographically diverse, represent different parts of the state, work in different practice areas, and come from a variety of organization types. Supervisors should be similarly diverse and should work in organizations that have demonstrated a commitment to diversity, equity, and inclusion.

6. Provide Training and Support for Supervisors and Examiners. Good supervisors and examiners will bring their own expertise to the licensing system. It is not necessary to erase differences of opinion among these experts or train them to reach agreement on every aspect of a candidate's portfolio. One of the strengths of a licensing system based on supervised practice is that it recognizes nuances and differences in approaches: These are the hallmarks of a profession.

Supervisors and examiners, however, will benefit from regular training on matters such as the purpose of the assessments used; the scoring of those assessments; methods to avoid implicit bias; approaches to providing constructive feedback; and ways to inform candidates that their work is not competent. It is also useful for examiners to meet periodically and review sample portfolios together; this can help individual examiners expand their perspectives. Supervisors may also appreciate a support network that allows them to share tips on providing constructive feedback and discuss other aspects of their role.

7. Make the System User Friendly. Training new lawyers, providing feedback, and rating performance take time, but a user-friendly system will greatly reduce that time commitment. System designers should solicit input from supervisors and new lawyers to understand the tools that will help users provide efficient, effective feedback. Some workplaces may have existing practices that will inform the new system. Best practices in performance review could help in implementing the system.

Supervisors and candidates often find online systems convenient. These systems also allow for ready sharing of materials with examiners. Some fields have even developed apps for providing feedback or assessments through smart phones. Designers, however, may prefer to use a paper

system during the early phase of implementation. This allows users to identify flaws and suggest improvements before putting the system online.

8. Design High-Quality Feedback and Assessment Tools. A good licensing system will use a variety of tools for feedback and assessment—just as the bar exam uses varied question formats. The literature offers these tips for designing fair and reliable tools that contribute to valid uses:

- Tie feedback and assessment to specific components of competence.
- Operationalize those components with prompts that remind supervisors of actions signifying levels of developing competence.
- Avoid numerical ratings on the forms: supervisors find them difficult to assign and err towards high ratings. Candidates also receive little guidance from numbers. Scores, however, may be built into the system for review by examiners.
- Encourage narrative feedback and ensure that the feedback is shared with the candidate.

The final pages of this handout show two sample feedback forms following this guidance. The forms are based on competencies identified by the *Building a Better Bar* report. The sample forms focus on a client interview; other forms would focus on different tasks and highlight different competencies.

The sample forms offer just two options for providing feedback and assessing competence. A candidate's final portfolio should include multiple feedback forms of different types, reflections from the candidate, and samples of the candidate's work product (with client-identifying information deleted).

9. Be Transparent. Transparency is essential to fairness. Candidates should know who will evaluate them, how they will be evaluated, and the criteria for successful performance. Transparency also enhances reliability and valid uses of assessment because the system is open to criticism and improvement.

10. Evaluate and Adjust. Assessment systems evolve as they benefit from experience and encounter new challenges. A good system will provide for regular evaluation and feedback from stakeholders. For a lawyer licensing system, those stakeholders include practicing lawyers and judges, legal educators, law students, recent graduates, clients, and the public at large. Regular assessment will assure that the licensing system retains credibility and protects the public.

Bibliography

David A. Cook et al., *When Assessment Data Are Words: Validity Evidence for Qualitative Educational Assessments*, 91 ACAD. MED. 1359 (2016).

David A. Cook et al., *A Contemporary Approach to Validity Arguments: A Practical Guide to Kane's Framework*, 49 MED. EDUC. 560 (2015).

Anthony A. Donato et al., *Validity and Feasibility of the Minicard Direct Observation Tool in 1 Training Program*, J. GRAD. MED. EDUC., June 2015, at 225.

Marjan Govaerts & Cees P.M. van der Vleuten, *Validity in Work-Based Assessment: Expanding Our Horizons*, 47 MED. EDUC. 1164 (2013).

Michael T. Kane, *Validating the Interpretations and Uses of Test Scores*, 50 J. EDUC. MEASUREMENT 1 (2013).

Michael Kane, Terence Crooks & Allan Cohen, *Validating Measures of Performance*, EDUC. MEASUREMENT: ISSUES & PRACTICE, Summer 1999, at 5.

Nyoli Valentine et al., *Fairness in Human Judgment in Assessment: A Hermeneutic Literature Review and Conceptual Framework*, 26 ADVANCES IN HEALTH SCIENCES EDUC. 713 (2021).

Nyoli Valentine et al., *Making It Fair: Learners' and Assessors' Perspectives of the Attributes of Fair Judgment*, 55 MEDICAL EDUC. 1056 (2021).

John J. Norcini & Danette W. McKinley, *Assessment Methods in Medical Education*, 23 TEACHING & TEACHER EDUC. 239 (2007).

Cees P.M. van der Vleuten & Lambert W.T. Schuwirth, *Assessing Professional Competence: From Methods to Programmes*, 39 MEDICAL EDUC. 309 (2005).

Cees P.M. van der Vleuten, *Revisiting "Assessing Professional Competence: From Methods to Programmes,"* 50 MEDICAL EDUC. 885 (2016).

John Q. Young & Matthew McClure, *Fast, Easy, and Good: Assessing Entrustable Professional Activities in Psychiatry Residents with a Mobile App*, 95 ACAD. MED. 1546 (2020).

7 The Fairness of Comparing Test Scores across Different Tests or Modes of Administration

Mary Pommerich[1]

Introduction

There is often a desire among test users to compare scores across tests of similar content that are developed by competing test publishers for similar purposes. In cases where appropriate data are available, statistical methods can be applied to link scores across the different tests, facilitating the comparison of scores. College admissions is one realm where this practice occurs regularly. Given that different tests have different characteristics, the question arises as to whether it is fair to compare scores that have been linked across different tests. A different, yet related, concern arises when users wish to compare scores across the same test that is given under different modes of administration. Under this scenario, the test publisher might apply statistical methods to link scores across the different modes of administration, facilitating the comparison of scores. A testing program converting from paper-and-pencil to computer administration is one realm where this practice can occur. Given that the test differs in terms of how it is administered, the question arises as to whether it is fair to compare scores across the different modes of administration.

Fairness in testing has been addressed extensively in the measurement literature. However, the scope of the discussion typically focuses on a single test and does not include the contexts described here. Although practitioners have argued that it is the use of a test (or test scores) rather than the test itself that is fair (Camilli, 1993, 2006; Darlington, 1971; Thorndike, 1971), fair use of *linked* scores is usually not considered outside of the context of alternate forms of the same test.[2] Linked scores are scores that have been statistically linked so as to enable identification of *concordant* or *comparable* scores across tests or modes of interest. While sources of mode effects and the comparability of scores across administration modes are fairness concerns that have been investigated thoroughly in the measurement literature, comparability studies tend to focus on comparing scores across modes of administration rather than the *fairness* of comparing scores across modes of administration.[3] A subtle distinction perhaps, but important. Fairness may be implicitly assumed where score comparability is deemed to hold or scores have been linked across modes, but is that really the case? Considering score comparability from a fairness perspective may change how results are viewed. Likewise, while the limitations of linking scores across distinct tests have been addressed to some degree in the measurement literature (e.g., Dorans, Pommerich, & Holland, 2007; Feuer, Holland, Green, Bertenthal, & Hemphill, 1999; Pommerich & Dorans, 2004), the fairness of comparing scores that have been linked across different tests has not been a focus of the linkage literature.[4] This may be because fairness is a complex concept with social implications that make it difficult to address.

DOI: 10.4324/9781315774527-9

Although it may appear that comparing scores across different modes of administration and comparing scores across different tests are fundamentally different issues, there are some important commonalities to the two scenarios pertaining to score linking. First, the same statistical methods are often used across the two scenarios to link the scores to be compared. Second, linked scores may be substituted for actual scores on the test/mode not taken, and used to make decisions. Third, users may be inclined to treat linked scores as if they can be used interchangeably across tests/modes when such an interpretation may not be warranted. If this is the case, the use of linked scores can result in decisions that differentially impact individuals and/or groups. As Feuer et al. (1999) noted, test-based decisions involve error, and linkage can add to that error. Making decisions based on scores that have been linked across different administration modes or different tests creates possibilities for unfairness above and beyond that associated with making decisions based on scores from a single test or mode. Hence, it makes sense to discuss these two scenarios together.

This chapter focuses on two primary questions:

- Is it fair to compare scores that have been linked across different modes of administration?
- Is it fair to compare scores that have been linked across different tests?

First, a fairness overview is given and related to the context of interest. Next, a linking overview is given and likewise related to the context of interest. Specific fairness issues pertaining to comparing linked scores across different modes of administration and comparing linked scores across different tests are then elaborated upon.

Fairness Overview

In considering fairness issues associated with using linked scores, it appears essential to start with a working definition of fairness, to provide a setting for the discussion to follow. Defining fairness is a tricky problem, however, as there is no definition that is generally accepted by all (Cole & Zieky, 2001; Zieky, 2006). What constitutes fairness can be viewed as a social question (Willingham & Cole, 1997) with judgments of fairness driven by values that are likely to differ across people (Sawyer, Cole, & Cole, 1976). Darlington (1971) concluded that the term "fair" carries various connotations that generally conflict with each other and that no single test is likely to meet all the requirements needed for a fair test. The *Standards for Educational and Psychological Testing* similarly state that the term fairness has no single technical meaning and outline four general views of fairness: as equitable treatment in the testing process, as lack of bias, as access to the construct(s) measured, and as validity of individual test score interpretations for the intended uses (AERA et al., 2014).

In spite of the elusiveness of the concept of fairness, a couple of general fairness perspectives do stand out in the literature as being relevant to the current context. The first perspective suggests that fairness is a property of test use rather than the test itself (Camilli, 1993, 2006), and that a test may be fair for some uses but not others (Darlington, 1971; Thorndike, 1971). Although this perspective is not presented in terms of validity, test use is inherently associated with validity (Kane, 2013). The second perspective explicitly argues that the most meaningful definition of fairness is based on validity, because anything that lowers the validity of a test for a group reduces the fairness of the

test (Zieky, 2006). More specifically, fairness is defined as *comparable validity* for individuals and groups at each assessment stage (Willingham & Cole, 1997; Xi, 2010). This perspective is related to the test use perspective in that, if score-based inferences are not equally valid for all relevant groups, decisions derived from those score inferences will not be fair (Langenfeld, 1997). Unfortunately, tying fairness to validity does not appear to provide a means for establishing a generally accepted definition of fairness—because validity is a matter of degree that may be interpreted differently based on personal values, fairness remains a matter of degree too (Cole & Zieky, 2001).

Different Modes Example

Both of the perspectives addressed above are relevant to the questions at hand. The issue lies with using linked scores in place of actual scores to make decisions about examinees. Consider the case where a test that has historically been administered via paper-and-pencil (P&P) has been converted to a computerized administration. This scenario is depicted in the left side of Figure 7.1. The test has likely been studied extensively with regard to validity for the P&P mode but not for the computer mode, and decision/selection criteria for test users will likely have been established based on the P&P test. Thus, fairness in terms of decision/selection outcomes will likely have been evaluated within the context of P&P administration. Green (1984, p. 77) stated the concern well for this type of scenario: "When a conventional test is transferred to the computer, it brings its validity with it. At least, we hope it does."

If scores from a computer administration have been linked to scores from a P&P administration to identify "comparable" or "concordant" score points across the two modes, decisions for examinees that take the test via computer will be based on concordant P&P scores rather than actual P&P scores. Note that comparable scores are not the same as interchangeable scores (to be addressed in more detail in the linking overview). Interchangeable scores are the ideal outcome of a linkage, while comparable scores imply a lower level of association (i.e., the linked scores can be compared across modes but not treated interchangeably). Interchangeable scores are expected when scores are equated across alternate forms of a test meeting certain prerequisites, but not necessarily when scores are linked across alternate modes of administration. Alternately, it might not be necessary to link scores across modes if evidence suggests they can be treated interchangeably without adjustment. Drasgow and Chuah (2005) advised that if a computerized administration does not yield scores that are equivalent to scores from a P&P administration, the test must be revalidated for the computer mode. Likewise, testing programs that choose not to revalidate should show strong evidence that scores are equivalent across modes.

The term *score equivalence* has been used to signify different things in the mode effects literature. The American Psychological Association (1986) stated that scores across modes of administration may be considered equivalent when score distributions are approximately the same across modes and individuals are rank ordered in approximately the same way. Elsewhere, the term score equivalence has been used to describe a situation where score distributions are approximately the same (e.g., Lottridge, Nicewander, Schulz, & Mitzel, 2008). In an attempt to clarify the terminology, the term *distributional equivalence* will be used here to describe the situation where score distributions are the same across two modes. It is generally expected that distributional equivalence should hold when comparing scores across modes (e.g., Kolen, 1999; Lottridge et al., 2008; Wang & Kolen, 2001).

Beyond distributional equivalence, researchers have suggested the need for *construct equivalence* across modes (e.g., Eignor, 2007; Kolen, 1999; Lottridge et al., 2008; Sawaki, 2001), such that the construct being measured across the two modes is equivalent. The presence of distributional equivalence and construct equivalence would be consistent with the APA (1986) definition of score equivalence (i.e., similar rank ordering of scores across modes is evidence of construct equivalence). However, the APA definition of score equivalence is less rigorous than that for the classical test theory definition of parallel tests, where equal correlations with criterion variables would also be needed for scores to be treated interchangeably across tests or modes (Bugbee, 1996). In the case of non-parallel tests, equal predictions of external criteria would not be attained because reliabilities are unequal (Neuman & Baydoun, 1998).

Hence, some researchers have suggested the need for *predictive equivalence* across modes (e.g., Bugbee, 1996; Neuman & Baydoun, 1998; Wolfe, Moreno, & Segall, 1997), such that external criteria are predicted equivalently, while others have suggested the need for *correlational equivalence* across modes (Bugbee, 1996; Kolen, 1999; Zitny, Halama, Jelinek, & Kveton, 2012), such that scores correlate equivalently with external criteria. The comparable validity perspective of fairness espoused by Cole and Zieky (2001) and Xi (2010) suggests that distributional equivalence, construct equivalence, and predictive equivalence[5] might all be needed to be truly fair when comparing scores across different modes of administration.

Different Tests Example

The case where scores have been linked across two distinct tests presents a similar concern to the different modes example, in that each test has likely been validated extensively for the particular uses that are specific to them (such as making selection decisions), but the use of linked scores as a substitute for actual scores has likely never been validated. Take the realm of college admissions. ACT and/or SAT scores are used by many post-secondary institutions in their admissions process. Although the market is changing, the ACT has generally been more popular in the central United States, while the SAT has been more popular along the east and west coasts. Hence, schools are likely to set their selection criteria on the basis of the test that is dominant in their region. Because many schools now accept ACT or SAT scores, there is a desire for a linkage between scores on the two tests to ensure that comparable decisions are made regardless of the type of test scores submitted. (Alternatively, schools may choose to evaluate validity, fairness, and selection criteria for each test and maintain separate systems.) An institution that relies on a linkage may develop its own, or use one that has been provided by the test developers (e.g. Dorans, Lyu, Pommerich, & Houston, 1997).

Consider the scenario where a school that has historically used SAT scores in the admissions process (i.e., evaluated validity and fairness and set selection criteria based on the SAT) now also accepts ACT scores and uses a linkage to facilitate the decision-making process. This scenario is depicted in the right side of Figure 7.1. In this case, ACT scores would be linked to SAT scores and individuals submitting ACT scores would be assigned concordant SAT scores, and decisions would be made based on the concordant SAT scores rather than actual SAT scores. In this scenario, distributional equivalence is likely to hold across the two tests, but construct equivalence and predictive equivalence might not. This raises questions about the inherent fairness of comparing scores that have been linked across different tests.

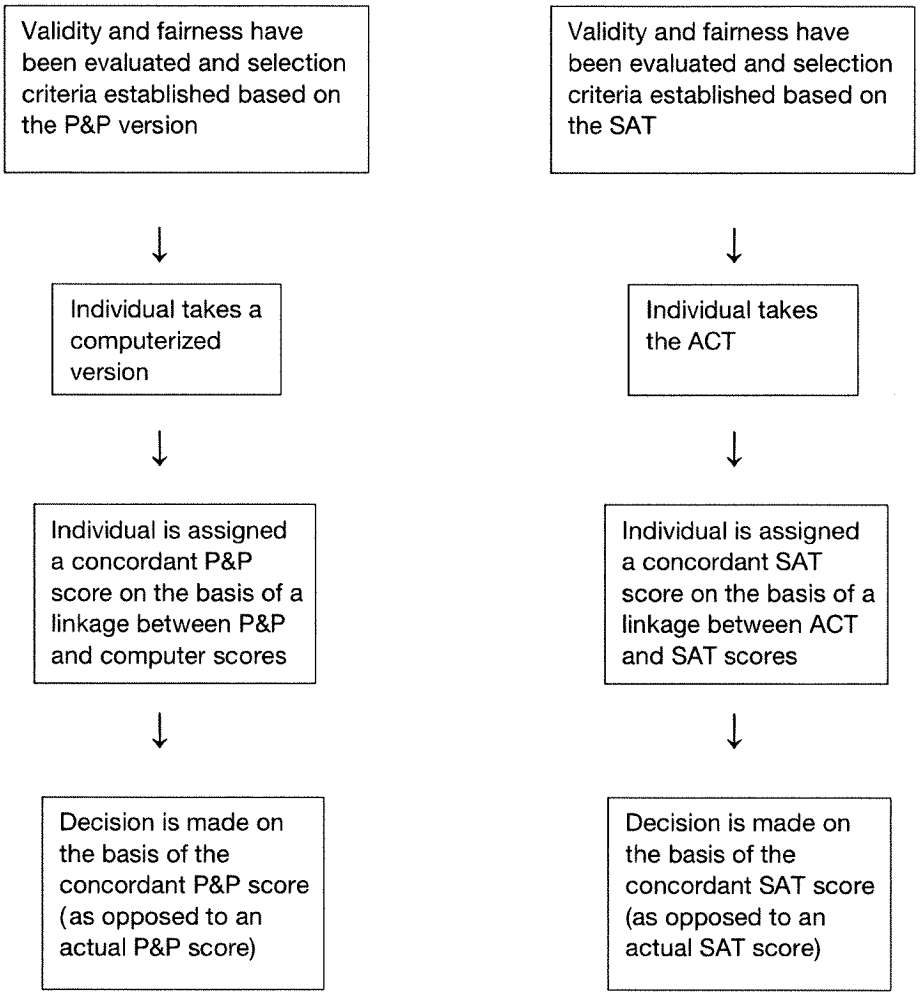

Figure 7.1 Linkage Scenarios for a School that has Historically Used P&P Scores to Make Decisions (Left) and for a School that has Historically Used SAT Scores to Make Decisions (Right)

Fairness to Individuals versus Fairness to Groups

In attempting to define fairness in the current context, it is important to make distinctions between fairness to individuals and fairness to groups because both outcomes might not be equally attainable. Thorndike (1971) demonstrated that there can be a trade-off between the two outcomes when making selection decisions (i.e., actions that are fair to individuals might be unfair to groups, and vice versa). Sawyer et al. (1976) framed this distinction in terms of maximization of success (based on individual parity) versus maximization of opportunity (based on group parity) and noted that the two concepts are often at odds, even though both are based on notions of individual merit. They also

noted that concerns about fairness (at that time) tended to require that selection be based on the merits of individuals without regard to their group membership. Those concerns may have shifted over time, as more recently Cole and Zieky (2001) bemoaned the fact that the study of individual differences has been overshadowed by the study of group differences. They noted that there is more individual variation of scores within groups than variation between groups, and suggested that all concerns of fairness for groups be applied to the issue of individual differences.

More recently, Camilli (2006) considered the differences between individual and group interpretations of fairness and concluded that the question of whether individuals are disadvantaged is not the same as the question of whether a group is disadvantaged, because the group question makes the assumption that individuals within each group are similar for the purpose of comparison. A related issue had been raised earlier by Breland and Ironson (1976), namely that the classification of individuals into groups is not necessarily readily achieved. Thus, fairness to individuals and fairness to groups are somewhat contrasting outcomes that might need to be considered separately.

Summary

In light of the above discussion, it appears that fairness is a rather nebulous concept, and that establishing a working definition of fairness in the current context is not an easy task! It does seem clear, however, that there is a relationship between fairness, validity, and score comparability. The mode effects literature suggests three desirable properties of fairness that would ideally be established when comparing scores that have been linked across modes of administration or different tests:

1. distributional equivalence (e.g., the score distributions are the same for the linked scores);
2. construct equivalence (e.g., the tests or modes measure the same construct to the same degree);
3. predictive equivalence (e.g., the tests or modes have the same predictive relationship with a criterion measure).

The predictive equivalence requirement is particularly pertinent from a fairness perspective because historically, a regression model approach has been widely used to evaluate fairness in selection (Cleary, 1968; Petersen & Novick, 1976), essentially examining whether the regression of the criterion onto the predictor space is invariant across groups (Dorans, 2004b).

Relatedly, an empirical measure of construct equivalence is whether the linking relationship is invariant across groups (Dorans & Holland, 2000). When invariance does not hold for a linking, the question arises as to whether you would make the same decisions using a linkage based on the total group versus using group specific linkages. Note that this is a different sort of concern regarding fairness to groups than expressed in the fairness literature, which has focused on differences in proportions selected across groups (e.g., Hartigan & Wigdor, 1989). Misclassification of individuals is also a concern when using linked scores to make decisions.

Linking Overview

Before proceeding with a more detailed discussion of the two scenarios of interest, comparing scores that have been linked across different modes of administration and

comparing scores that have been linked across different tests, it is helpful to provide some background on linking and define some of the relevant terms that are being used. Linking frameworks have been discussed in a variety of sources, including Flanagan (1951), Angoff (1971), Mislevy (1992), Linn (1993), Feuer et al. (1999), Dorans (2004a), and Kolen and Brennan (2014). Kolen (2004) compared and contrasted the various linking frameworks that have been defined. More recently, Holland (2007) and Holland and Dorans (2006) presented a linking framework that builds on the preceding frameworks; this chapter utilizes their framework and terminology.

For two forms (or modes of administration, or tests), a *link* between their scores is a transformation from a score on one to a score on the other. *Linking* is the means by which that transformation is obtained. Two categories of linking methods are germane in the current context: scale aligning and equating. *Scale aligning* has the goal of *comparable scores*, while *equating* has the more stringent goal of *interchangeable scores*: in the words of Dorans (2013), comparable scores are a necessary but not sufficient condition for producing interchangeable scores. Comparable scores have historically been defined as scores from tests with different psychological functions that are scaled to have the same distributions with respect to a particular group of examinees, with comparability assured only for that specific group taking the tests under specific conditions (Angoff, 1971).[6] Interchangeable scores are expected to have the same meaning across the two forms (or modes of administration, or tests), and so can be treated interchangeably. The *equipercentile function* is commonly employed in both scale aligning and equating. The equipercentile function links a score on Test X to the corresponding score on Test Y that has the same percentile in a target population. If the influence of the target population is small (i.e., the same results are likely to be obtained regardless of the population used to compute the linking function), the results are said to be *population invariant* (Dorans & Holland, 2000) and the linked scores are considered to be interchangeable. If there is a non-negligible influence for the target population, the linked scores are said to be comparable but not interchangeable.

A hallmark of equating is the rigorous requirements placed on forms/modes/tests to be linked. Dorans and Holland (2000) identified five requirements for a linking to be an equating: the tests should measure the same constructs (Equal Constructs Requirement) and have the same reliability (Equal Reliability Requirement), the function for linking the scores of Test Y to those of Test X should be the inverse of the function for linking the scores of Test X to those of Test Y (Symmetry Requirement), it should be a matter of indifference to an examinee to be tested by either one of the two tests that have been linked (Equity Requirement), and the choice of (sub)population used to estimate the linking function between the scores of Tests X and Y should not matter (Population Invariance Requirement). Additional requirements have been suggested by Kolen and Brennan (2014). The rigor of equating comes not from the statistical procedures applied to link the scores, but from the way the tests are constructed, namely to the same specifications (Mislevy, 1992), and from careful design of equating studies. This rigor is needed to ensure fair treatment of examinees, through the achievement of interchangeable scores.

Within the scale-aligning category of linking, there are two types of scaling that are pertinent to our fairness discussion: concordance and calibration. These and other types of scaling are delineated in Holland (2007) and Holland and Dorans (2006). The term *concordance* is assigned to a linking between forms/modes/tests that measure similar constructs at a similar level of reliability, while the term *calibration*[7] is assigned to a linking between forms/modes/tests that measure similar constructs but at a dissimilar

level of reliability. Concordance also assumes similar difficulty and similar populations across the tests being linked. Dorans et al. (1997) labeled their linkage between ACT and SAT I scores as a concordance. Eignor (2007) labeled linkages between P&P and computer adaptive test (CAT) scores as a calibration because the Equity Requirement of equating is not met, but suggested that linkages between P&P and computer-based test (CBT) scores (i.e., scores from a linear administration of a P&P test on computer) could be labeled an equating. Eignor also noted that calibrated scores are often treated as though they are interchangeable and questioned the appropriateness of doing so, and that concorded scores cannot be treated as interchangeable. This is in contrast with equating, which is intended to produce interchangeable scores because of its stringent requirements. These distinctions will be discussed in more detail later.

Practitioners may have differing viewpoints on what label to apply to a linkage and its outcome. For example, based on considerations of population invariance, Dorans and Holland (2000) and Dorans (2004a) suggested that a linkage (concordance) between ACT and SAT I math scores might yield nearly interchangeable scores, even though a commonly stated equating requirement was not met (i.e., the tests are built to different specifications). Contrary to Eignor's (2007) viewpoint, Schaeffer, Steffen, Golub-Smith, Mills, and Durso (1995) reported that they expected that GRE CAT scores would be interchangeable with scores earned on the P&P and CBT versions. Hence, we shouldn't assume that scores are or are not interchangeable on the basis of whether or not a linkage scenario meets all of the requirements viewed as necessary to be considered an equating. Proper interpretation of a linkage outcome (i.e., whether the linked scores are best viewed as comparable versus interchangeable) can depend on a number of factors, including linking methodology, design of the linking study, characteristics of the groups and tests being linked, and how linkage results will be used. There are some tools that can be used to evaluate the feasibility of treating linked scores as interchangeable, to be discussed later. When linked scores are comparable (i.e., score distributions are aligned) but not interchangeable across different tests or modes of administration, there is a potential for unfairness to individuals or groups because assigned scores, score meanings, and decisions made from these scores could vary depending upon which test or mode is taken, or which group an examinee belongs to.

Fairness Properties Revisited

With this in mind, it is helpful to tie the discussion of comparable scores and interchangeable scores back to the desirable fairness properties outlined earlier. Interchangeable scores as defined within an equating context (i.e., with regard to the equating requirements delineated by Dorans & Holland, 2000) appear to encompass the fairness properties of distributional equivalence, construct equivalence, and predictive equivalence, whereas comparable scores encompass distributional equivalence only. Hanson, Harris, Pommerich, Sconing, and Yi (2001) warned that it is possible to develop a link function that results in almost perfect comparability of distributions in one population, no matter how incomparable the two scores are for individuals. Eignor (2007) made a similar distinction between equivalent scores and scores that are equivalent in appearances only, noting that sets of scores that are identical in appearance share the same means, variances, and distributions of scores, but the scores themselves do not convey the same meaning. For these types of reasons, Lottridge et al. (2008) addressed the need to consider both distributional and construct equivalence when evaluating comparability across P&P and computerized tests. Dorans (2004b) suggested that three aspects of fairness should be

addressed by testing programs: population invariance in linking functions, differential item functioning, and differential prediction. This supports the notion that distributional equivalence, construct equivalence, and predictive equivalence are all needed for optimal fairness when comparing scores that have been linked across different modes or tests.

More Fairness Considerations for Comparing Test Scores across Different Modes

Now that sufficient background information has been provided for the context of interest (fairness issues with regard to using linked scores), the issue of comparing scores across different modes of administration can be discussed in more detail. Fairness is a concern when there are two modes of administration that are in concurrent use, or when there is a single mode of administration in use (such as computer) but scores are compared with scores from a prior mode of administration (such as P&P). Within the realm of computer administration, distinctions also need to be made for the administration algorithm (linear versus adaptive), and the delivery method (Internet versus local). Using the acronyms introduced earlier, CBT corresponds to a linear administration while CAT corresponds to an adaptive administration. Under linear administration, a fixed-form test is administered in a non-adaptive manner. Under adaptive administration, the test is tailored to each examinee, with items selected to adapt to the examinee's estimated ability. Under Internet delivery, the test is delivered over the Internet. Under local delivery, the test is delivered over a local network or on a personal computer. Concerns specific to Internet delivery will be considered later.

Extensive research has been conducted comparing performance across computer and P&P modes of administration. Wang, Jiao, Young, Brooks, and Olson (2008) noted that there were over 300 mode of administration effects (mode effects) studies conducted in 25 years, spanning the realms of intelligence, aptitude, ability, vocational interest, personality, and achievement tests. That number is likely to have increased in the interim. It is not the intent of this chapter to summarize all of the mode effects literature; readers are referred to Blazer (2010), Lottridge et al. (2008), Texas Education Agency (2008), and Paek (2005) for some recent, thorough reviews of mode effects research and findings in the realm of educational testing. In addition, it is not the intent of this chapter to address how to evaluate score comparability across modes of administration; readers are referred to sources such as Kolen (1999), Wang and Kolen (2001), Eignor (2007), Lottridge et al. (2008), Texas Educational Agency (2008), Karkee, Kim, and Fatica (2010), Schroeders and Wilhelm (2011), Chua (2012), Randall, Sireci, Li, and Kaira (2012), and Mroch, Li, and Thompson (2015) for various discussions about how to collect data and evaluate score comparability.

Mode Effects Research

In general, the reviews of the mode effects literature suggest that scores tend to be comparable across P&P and computer administrations more often than not. Paek (2005) asserted that sufficient evidence exists to conclude that computer administration does not significantly affect student performance, with the exception of tests containing lengthy reading passages. Blazer (2010) noted that there are very few differences in test scores for multiple choice tests across computer and P&P administrations, but cautioned that examinees' demographic characteristics and computer skills, computer and test characteristics, item type, and content area could all affect comparability. The Texas

Education Agency (2008) raised the question of whether enough evidence has been collected to determine that mode effects studies are no longer needed, but concluded that states need to assess their own situation and weigh the costs and risks of conducting/not conducting comparability studies. Recent meta-analyses of math tests (Wang, Jiao, Young, Brooks, & Olson, 2007) and reading tests (Wang et al., 2008) support the notion that comparability is more likely to be found than not, as did earlier meta-analyses (Bergstrom, 1992; Mead & Drasgow, 1993).

Although the overall trend may favor comparability, findings for specific tests may vary on an individual basis. Researchers have identified a number of concerns that need to be considered when comparing scores from P&P administration to computer administration. Kolen (1999) discussed a number of potential threats to score comparability across modes, including differences in test questions, differences in scoring, differences in operational testing conditions, differences in examinee groups, and violations of statistical assumptions for establishing comparability. Huff and Sireci (2001) elaborated on a number of potential threats to validity in computerized testing, including construct underrepresentation, construct-irrelevant variance, improper estimates of examinee scores, and unintended consequences. These concerns should not be ignored, as threats to score comparability and validity are also threats to fairness. Threats to score comparability could be addressed by linking scores across modes, but as discussed previously, that wouldn't necessarily ensure interchangeable scores.

In particular, differential access to computers is a notable fairness concern associated with mode of administration, related to socioeconomic status. If there is an advantage for taking a test on the computer rather than via P&P administration, then those examinees with less access to computers (and hence, potentially less familiarity with computers) could be disadvantaged. Concerns about a digital divide have been commonly raised, recently with regard to access to the Internet (Bartram, 2006). The U.S. Census Bureau (2014) reported that in 2012, 78.9% of all U.S. households had a computer at home, with 94.8% of those households using the computer to access the Internet, while overall, 74.8% of all U.S. households had Internet use at home. These computer/Internet usage statistics suggest that access issues are less of a concern now than in the past when computers and the Internet were more of a novelty, but that there still could be a digital divide that could threaten the fairness of comparing scores across modes of administration, especially for individuals with a lower socioeconomic status.

Linkage Issues

If a testing program conducts mode effects studies and finds that scores are not comparable across differing modes of administration, two approaches are commonly chosen. The first approach is to make iterative changes to the computer interface or administration in an attempt to eliminate mode effects. An iterative approach to computer interface development was demonstrated by Mazzeo, Druesne, Raffeld, Checketts, and Muhlstein (1991) and Pommerich (2004). The second approach is to link scores across the modes of administrations so that they can be compared. Eignor (2007) described in-depth ways one might design linking studies to relate scores across computer and P&P administrations, taking into consideration adaptive versus linear algorithms (i.e., CBT vs. CAT). Interchangeability of scores was a central concern in his discussion, particularly with regard to linkages between CAT and P&P scores. Rudner (2010) discussed a linkage study intended to equate CAT and P&P scores on the GMAT (i.e., yield interchangeable scores) and concluded that the CAT-based scaled scores were not truly equivalent to the P&P

scores even though the CAT scores were forced to the P&P scale. His experience demonstrates that merely conducting a linkage between scores across modes does not ensure that the desired outcome will be obtained and highlights the importance of evaluating the quality of a linkage, as recommended in Pommerich, Hanson, Harris, and Sconing (2004). When analyses suggest that linked scores are comparable but not interchangeable, then fairness is more likely to be a concern for reasons discussed earlier.

If a testing program chooses not to conduct mode effects studies or score linkages and compares scores across modes anyway, this would be similar to a *presumed linking* scenario, to borrow the terminology of Dorans and Middleton (2012). In a presumed linking scenario, comparisons of scores are made even though there is no evidence to support making them. There are some obvious fairness concerns in such a situation, as the degree of comparability in scores would be unknown, and individuals or groups could be negatively impacted by the fact that scores from one mode might not have the same value or meaning as scores from the other mode. In such a situation, Dorans and Middleton (2012) would recommend evaluating invariance relationships across the modes of administration to provide support for such a practice.

If mode effects studies are conducted and show evidence of no mode effects, testing programs may choose not to link scores across modes of administration. This is different from the presumed linking scenario, as there is evidence to support the decision not to link scores. Attention must still be paid, however, as to whether scores can be treated interchangeably across modes or not. Likewise, any time a linkage is conducted, there are fairness issues if comparable scores are treated as if they are interchangeable when they are not. If the name is any indication, comparability studies might be content to obtain comparable scores across linked modes, even though interchangeable scores would be fairer.

As discussed earlier, demonstrating distributional equivalence is sufficient evidence for comparable scores, but not necessarily for interchangeable scores. Lottridge et al. (2008) discussed the importance of evaluating construct equivalence in mode effects studies, how one might address comparability using a hypothesis-testing approach in a construct validation framework, and highlighted a number of mode effects studies that looked at various aspects of construct equivalence. Demonstrating construct equivalence as well as distributional equivalence of linked scores would be more in line with establishing interchangeable scores rather than merely comparable scores. On the other end of the spectrum, Winter (2010) asked the question of how comparable is comparable enough, and concluded that it depends on how the scores will be interpreted and used. She presented a continuum of score comparability that showed less score comparability required for pass/fail scores and achievement level scores than for scale scores and raw scores. Mroch et al. (2015) similarly suggested that score comparability is on a continuum between interchangeable and incomparable, where the required level of comparability is tied to how scores will be interpreted for a particular use. An evaluation of score comparability with regard to usage can be found in Kapoor and Welch (2011), who addressed the impact of mode of administration on proficiency classifications.

CAT Considerations

In considering the trend across mode effects studies favoring conclusions of comparability, there is one caveat that should probably be made. Namely, much of the research may have focused on CBT administration, which is a more straightforward (and more

common) means of administration than CAT. Recall Eignor's (2007) suggestion that linkages between P&P and CBT scores could be labeled an equating, indicating an expectation of interchangeable scores across the two modes. In contrast, a number of researchers have expressed serious concerns about the interchangeability of CAT and P&P scores (Eignor, 2007; Kolen, 1999; Wang & Kolen, 2001; Wang & Shin, 2010). Other researchers have expressed related concerns pertaining to the impact of item calibration medium (P&P or computer) on CAT scores (Choi & Tinkler, 2002; Pommerich, 2007a).

The primary concern for comparing scores that have been linked across CAT and P&P administration modes focuses on the Equity Requirement of equating and the fact that it is not likely to hold, given that CAT differs substantially from P&P in terms of items administered, administration conditions, and scoring methods. In particular, the Equity Requirement will not hold if tests have different conditional standard errors of measurement across CAT and P&P administrations. Equipercentile methods can readily be applied to link scores across the administration modes; however, the scores can still differ in their conditional precision across modes. Hence, some would say that CAT scores cannot be treated interchangeably with P&P scores because they differ in their statistical specifications (e.g., Eignor, 2007). The same argument could be applied to linkages between CAT scores and scores from a linear CBT.

Kolen (1999) also warned that sufficient differences could exist between CAT and P&P tests, such that the construct measured could be affected and various subgroups might favor one mode over the other. If examinee preferences do exist across administration modes for an operational test, that would be a fairness concern. Because of the inherent differences between CAT and P&P administration, Wang and Kolen (2001) recommended that comparability be carefully established and evaluated. The thorough research and evaluation conducted by the ASVAB testing program prior to implementing CAT administration (Sands, Waters, & McBride, 1997) represents an example of the level of consideration that might be needed to alleviate fairness concerns, especially if CAT and P&P scores are to be used interchangeably.

Alternate Takes on Mode of Administration

There are a couple of variations on mode of administration that should be considered also with regard to score comparability. The first variation pertains to Internet delivery of computerized tests. The comparability between P&P and Internet administrations has not received a lot of attention in the realm of mode effects (Baumer, Roded, & Gafni, 2009; Naglieri et al., 2004). The second variation pertains to the use of unproctored Internet administration.

Internet Delivery

A notable concern associated with Internet delivery (also referred to as online testing) is that equipment and/or configurations can vary across administration locations, resulting in a loss of standardization (Bennett, 2003; Bridgeman, Lennon, & Jackenthal, 2003). Dorans (2012) raised fairness concerns associated with a loss of standardization in testing due to the increased use of technology. Naglieri et al. (2004) advised that the effects of mode of administration and the delivery method should both be studied to ensure the appropriate use of tests on the Internet. If a testing program utilizes Internet delivery of a computerized test in conjunction with local delivery or P&P administration, but doesn't

study its impact (or link scores) across the different modes/delivery approaches, this invokes the fairness concerns discussed earlier for a presumed linking scenario.

Bennett (2003) highlighted a critical concern about Internet delivery of tests, discussing ways in which the Internet connection could cause item presentation to vary across equipment, focusing on possible sources of delays in the administration of items. Significant delays and/or problems in Internet delivery of several statewide assessments made the national news in the spring of 2013. In Minnesota, examinees experienced computer slowdowns, freezes, and other problems. In Indiana, some examinees were locked out of the testing website during their exams, while others were unable to log into their exams at all. In Kentucky, school districts reported slow and dropped Internet connections, resulting in the temporary suspension of online testing. In Oklahoma, servers crashed, preventing examinees from completing their tests. In prior years, similar problems were noted with online testing in Wyoming, Virginia, and Texas. Wyoming's experience caused the state to abandon online testing and revert back to P&P administration. Across all of these states, the frequency and magnitude of the problems observed suggested that the vendors providing Internet delivery were not prepared to handle the kinds of demands that statewide administration placed on the test delivery system.

Clearly, the types of problems that can occur with Internet delivery of statewide assessments are a major threat to validity and raise many fairness concerns. Validity questions include whether construct-irrelevant variance is introduced for examinees with disrupted sessions. Namely, is the test measuring the same construct across examinees that are and are not affected by Internet delivery problems? Fairness questions include whether all students are equitably treated in the testing process (AERA et al., 2014) and whether scores are interchangeable across disrupted and non-disrupted sessions. Analyses conducted by independent parties for Minnesota and Indiana in 2013 concluded that the Internet delivery problems did not affect performance. In Indiana, this conclusion was drawn on the basis of the fact that examinees that were interrupted had gains across years as high as examinees that were not disrupted (Stokes, 2013). While this may be true for the group of interrupted students as a whole, there are likely to have been individuals that did not experience gains across years as a result of the disruption. Given the high-stakes nature of the assessments, there was also likely a negative psychological impact experienced by examinees whose sessions were disrupted.

Unproctored Administration

A relatively new approach to test administration that has been broached primarily for use in employment testing is unproctored Internet administration (Tippins, 2009; Tippins et al., 2006). Under this approach, examinees first take an unproctored test via the Internet and then take a shorter, proctored test to verify that their unproctored scores represent their abilities. Such an approach would allow greater flexibility about where initial testing takes place and could reduce costs, but also introduces a number of fairness concerns. Clearly, cheating would be a key concern with unproctored administration. Cizek (1999) stated that nearly every research report on cheating has concluded that cheating is rampant. However, Drasgow, Nye, Guo, and Tay (2009) noted some recent studies that indicated that cheating on unproctored tests may not be as widespread as thought for some types of tests. Verification testing is intended to detect cheaters, but if large numbers of examinees do cheat on the unproctored version, validity of the unproctored form is threatened, and there might be little cost savings to using unproctored

assessment. Loss of standardization due to the variation in equipment is also a potential concern in this scenario, as discussed earlier with regard to Internet delivery. Likewise, there would be concerns about the comparability of scores across unproctored and proctored settings. The concerns about a presumed linking scenario discussed previously would be relevant here too.

Future Considerations

In the future, we could see technology utilized in testing in ways that could result in an even greater loss of standardization. Pommerich (2012) noted that items and test characteristics can vary across examinees when CAT administration is used, equipment can vary across examinees when Internet delivery is used (i.e., if a testing program uses readily available equipment), and environment can vary across examinees when unproctored administration is used (i.e., if examinees test in their own home). It is possible that we could reach the point where no two examinees take a test under the same conditions. Given the prolific use of smartphones (and more recently, tablets), the profession will likely need to adapt to new ways of presenting and responding to tests that have not yet been extensively studied. Who knows what else the future may hold that could introduce even greater change into how people take tests—examinees taking tests while riding in self-driving cars or drones, using in-vehicle communication systems, perhaps? Dorans (2012) expressed a concern that the measurement profession has "lost sight of the essential need for controlled conditions of measurement" and gave examples associated with technology-driven assessment. His examples emphasize the importance of taking active steps to address the limitations of technology-based assessment and adapt our practices to compensate for them (Pommerich, 2012). Any changes to mode of administration should raise fairness concerns along the lines discussed in this chapter, until sufficient research has been conducted to alleviate them.

More Fairness Considerations for Comparing Test Scores across Different Tests

The issue of comparing scores across different tests is a more extreme scenario than comparing scores across different modes of administration for the same test. Under this scenario, scores from two distinct tests that are built to different specifications and administered to different populations are compared. It is more extreme than comparing scores across different modes of administration because the two tests are typically separate entities that are developed independently by different parties. These tests are likely not developed with any intention of linking scores across the two tests. Further, the linkages are typically conducted using a convenience sample of examinees that have taken both tests rather than based on a formal data collection design. If linked scores are used in place of actual scores to make decisions (Figure 7.1), this is a specialized type of test use that is probably not validated, and the test that is used to assign a concordant score is only indirectly being used for its intended purpose.

The discussion in this section focuses on a concordance situation, where the tests measure similar constructs, have similar levels of reliability and difficulty, and are administered to a similar population, and focuses on the equipercentile method as the means of linking scores. Concordances between scores from two college admissions exams, ACT and SAT, will be used as an example throughout the discussion to provide a familiar context from which to address fairness issues. The discussion here does not consider a presumed linking scenario

(Dorans & Middleton, 2012), as simple comparisons of ACT and SAT percentiles will lead to incorrect inferences about the relative performance of examinees on the ACT and SAT, due to population differences (Dorans & Petersen, 2010).

A limitation of concordances is that they don't typically result in scores that can be used interchangeably (with possible exceptions, such as noted by Dorans, 2004a and Dorans & Holland, 2000, discussed earlier). Unfortunately, evidence suggests that users are inclined to treat concordant ACT-SAT scores as if they are interchangeable or as predictions of scores on the test not taken. Thus, by virtue of making concordance results available, we create the potential for misuse and misinterpretation of those results, which raises concerns about fair treatment where concordant scores are used to assist in making selection decisions. Lindquist (1964) argued against creating ACT-SAT concordances because of concerns about misuse and misinterpretation; these concerns have not been alleviated over time. Pommerich et al. (2004) demonstrated that using equipercentile results for different purposes from which they are intended could give very misleading results for some examinees. They cautioned that if equipercentile-based concordant scores are used as a prediction of an individual's score (a misuse), the consequences should be considered, since the concordant scores will deviate to some degree from what the actual scores would be. Equipercentile-based linkages will be fair to the group used to conduct the linkage, in that equal percentages from the group will be selected using either test at concordant score points, but they will not necessarily be fair to individuals or specific subgroups, or to the larger population in which they will be applied (i.e., examinees taking the ACT or the SAT, but not both).

Brennan (2007) maintained that arguing against using comparable scores as if they were interchangeable might be a lost cause, but that cautioning users about potential errors in doing so is both necessary and possible. This is a call for disclosure that brings to mind Cole and Zieky's (2001) fairness recommendation that the measurement community take a greater leadership role in educating the public about potential misinterpretations of group differences (not pertaining to concordances) by addressing them directly in test materials and public discussions. They noted that the new faces of fairness require measurement professionals to react more directly and forcefully against instances of test misuse. The importance of disclosure or public education in a concordance scenario should not be underestimated. Pommerich (2007b) proposed five goals to strive for when conducting concordance. The goals, labeled the **FRANK** goals, are modified here to represent all linkage scenarios:

1. **F**lexibility in linking practices;
2. **R**esponsibility in creating and disseminating linkage results;
3. **A**wareness of the limitations of linkages;
4. **N**otification as to proper interpretation and use of results;
5. **K**nowledge of users and their practices.

In devising her **FRANK** acronym, Pommerich inadvertently channeled the thinking of Cronbach (1980), as cited in Linn (1984, p. 45), who stated that "the more we learn, and the franker we are with ourselves and our clientele, the more valid the use of tests will become." Full disclosure will allow test users to make informed choices about how to use concordance results and to better understand what impact their use may have on fairness.

On the other hand, Sawyer (2007) expressed a more realistic[8] viewpoint about the use of ACT-SAT concordances, stating that there is a sense among users that in the big

scheme of things, ACT-SAT concordance tables are probably good enough for the uses they are put to. In some regards, he may be right, as concordant scores are not likely to be the only basis for an admissions decision. In addition, testing standards and guidelines indicate that some responsibility for proper interpretation and use of concordance results should lie with test users, not just test developers (AERA et al., 2014; Joint Committee on Testing Practices, 2004; NCME Ad Hoc Committee on the Development of a Code of Ethics, 1995). Feuer et al. (1999) stated that policymakers and educators must take responsibility for determining the degree to which they can tolerate imprecision in linking. This might be a bit of a catch-22, however, as many users might be unaware of the inherent problems with a linkage (i.e., lack of precision, lack of interchangeable scores) and fail to grasp the implications for decisions that are made on the basis of that linkage. Thus, the onus of disclosure and education appears to fall back onto the test developer or whoever conducts the linkage that is provided to users.

In Sawyer's (2007) perspective, a more pressing concern is that users might think that concordance between *any* two tests is unproblematic. Pommerich (2007b) noted that situations where concordance is not appropriate might be less apparent than situations where equating is not appropriate. Dorans (2004a) and Dorans and Walker (2007) proposed an index of reduction in uncertainty (RiU) to help decide whether to utilize concordance or prediction methods to link two sets of scores, and concluded that a correlation coefficient of 0.866 is needed between the scores on the two tests to reduce the uncertainty of knowing a person's score on one test by at least 50%, given the score on the other test. By this standard, concordance is not viewed as appropriate for tests where the correlation falls below 0.866. Fairness is already a concern for a concordance situation, and it becomes a greater concern when concordances are conducted between tests that do not correlate at that high level.

Assessing the Interchangeability of Linked Scores

Consistency rates (i.e. the percent of consistent classifications using concordant versus actual scores) provide evidence of the degree of misclassification that might be expected by using concordant scores in place of actual scores and serve as a means of approximating the departure from equity (Hanson et al., 2001). Pommerich et al. (2004) demonstrated how the disparity between actual and equipercentile concordant scores for examinees taking both the ACT and SAT I increased as the correlation between linked test scores decreased. They cautioned that although the equipercentile method will yield score points that result in the same percentages being selected on either test, the same individuals will not necessarily be selected. When two tests being linked are highly correlated, the consequences of using equipercentile results at an individual level should not be too severe. However, as the relationship between the two tests decreases and the consistency of classification based on concordant and actual scores lessens, it might be meaningless to use equipercentile concordances even at a group level, as intended (i.e., to set equivalent cutscores). Practically speaking, there may be little point to selecting equivalent proportions across the two groups if individuals would be classified differently by the two measures.

Population invariance of linking functions is another means by which to evaluate whether fairness is likely to be a concern for score linkages. When population invariance does not hold, linked scores are comparable but not interchangeable, and attempts to use scores interchangeably could result in unfair treatment of some examinees or groups. Recently, assessment of the invariance of linkings across important subpopulations[9] has

received considerable attention in the measurement literature as a tool to assess the degree of interchangeability of scores (e.g., Dorans, 2004c; von Davier & Liu, 2008). Huggins and Penfield (2012) noted that the criteria of population invariance in linking functions (also referred to as score equity assessment) is becoming well-established as a necessary condition for fairness in tests that employ any form of linking. Score equity assessment was introduced and placed in a fairness context in Dorans (2004b) as a means to assess the fairness of a "test score exchange process" (Dorans & Liu, 2009). Violations of population invariance are a threat to test fairness because examinees from different groups that have the same score on one test will have different linked scores on the corresponding test, resulting in potential disadvantages for some group members (Huggins & Penfield, 2012). Dorans (2004b) recommended that score equity assessment be routinely addressed as a fairness consideration,[10] along with differential item functioning and differential prediction. The instability of linkages over time is another form of lack of invariance (Thissen, 2007) with implications for fairness too. If the test-taking populations change over time, a given concordance relationship may no longer hold. To ensure stability over time, linkages should be updated frequently.[11]

Recent applications of score equity assessment focused on score linkages across variations of the same test (e.g., Liu, Cahn, & Dorans, 2006; Liu & Dorans, 2013). The same approach can be applied to scores linked across different modes of administration for the same test or to scores linked across distinct tests. For example, Liu, Dorans, and Moses (2010) used score equity assessment to evaluate the population sensitivity of the most recent ACT-SAT concordances. The expectation that concordances are unlikely to be population invariant (Dorans & Petersen, 2010) was upheld for some groups/concordances that demonstrated a "substantial degree" of subgroup sensitivity. Other groups/concordances showed results that were essentially invariant. Likewise, Yin, Brennan, and Kolen (2004) evaluated invariance of concordances between ACT and Iowa Tests of Educational Development scores and found population invariance was upheld for some tests/linking methods, but not others. Dorans and Holland (2000) recommended creating different concordances for important subgroups when concordance results deviate considerably from invariance. However, the use of different concordances across different gender/ethnic groups, although intended to be fair, could be viewed as unfair by an undiscerning public, because it means that scores would be treated differently across groups (Dorans, 2004a). While the measurement community may not agree as to what can be done about lack of invariance in a concordance situation, we would all benefit from a public discussion (i.e., full disclosure) of the issues and the implications for common uses of the concordances.

Concluding Comments

Fairness proponents advocate promoting fairness at all stages of assessment from conception through score usage (Cole & Zieky, 2001; Downing & Haladyna, 1996; Kunnan, 2000; Willingham & Cole, 1997; Zieky, 2006). If linked scores are to be used to make decisions about examinees, then that type of usage should ideally be accounted for in fairness planning and evaluation. Xi (2010) demonstrated how a fairness argument might be built and substantiated in the context of a validity argument. Her fairness argument included a series of rebuttals that might "challenge the comparability of scores, score interpretations, score-based decisions and consequences for sub-groups" on the TOEFL iBT. She noted that to substantiate the argument, evidence has to be obtained that supports the comparability of the score-based interpretations and uses for relevant

groups. It appears that the question of whether it is fair to compare scores that have been linked across different modes of administration could be readily incorporated into a fairness argument such as this for a specific test of interest.

Ideally, the question of whether it is fair to compare scores that have been linked across different tests would also be incorporated into a fairness argument (most likely pertaining to the utilization of test scores), but this is a more awkward proposition given that a test is usually developed and evaluated as an autonomous unit and fairness is usually addressed with regard to a single test. This brings us back to the question raised earlier of whose responsibility it is for the proper interpretation and use of linkage results, especially when the linkage involves two distinct tests. Although test users have a number of responsibilities pertaining to test score use (AERA et al., 2014; Joint Committee on Testing Practices, 2004; NCME Ad Hoc Committee on the Development of a Code of Ethics, 1995), it might seem logical to expect that if a test developer creates a linkage between scores on their test and an external test and disseminates that linkage to users, then the test developer should explicitly account for that type of usage in a fairness argument or framework. Conversely, if a user develops a linkage between two distinct tests independent of the test developers, then the responsibility of the fairness argument should, in theory, lie with the user. Unfortunately, test developers are not likely to develop a fairness argument for comparing scores that have been linked across different tests because it involves a test that falls outside of their scope of control and because aspects of the fairness argument (i.e., establishing predictive equivalence) would require the involvement of test users.

At the heart of the matter when using a score linkage to make decisions is the interchangeability of the linked scores (or lack of interchangeability of scores). In terms of the fairness properties outlined earlier, interchangeable scores imply distributional equivalence, construct equivalence, and predictive equivalence, while comparable scores imply distributional equivalence only. Fairness will always be threatened to some degree if analyses suggest linked scores are comparable but not interchangeable, but scores are used interchangeably anyway. However, in any linkage, even an equating, scores are not likely to be perfectly interchangeable. Liu and Walker (2007) maintained that the issue should be one of degree, namely whether requirements are met sufficiently such that scores can be treated as interchangeable, within a reasonable amount of error. Some tools were discussed earlier to evaluate whether scores can reasonably be treated as interchangeable.

If linking conditions are such that interchangeable or nearly interchangeable scores are not possible, it should be asked whether inappropriate or unfair decisions could be made or whether inappropriate or unfair conclusions could be drawn as a result. If the answer is yes, then the linkage might not be defensible. A conservative approach in such a case would be to maintain separate score scales for the two tests (or modes) in question, but that might not be palatable to policymakers. If it is necessary to report the linkage results, then it is of utmost importance to fully disclose all linkage details and educate users on proper and improper usage of the linked scores. Knowledge of user practices is helpful to the degree that consequences of misuse and/or misclassification can be taken into consideration when providing guidance. The amount of error in linked scores should be reported and explicitly discussed so that users can make an informed decision on whether and how to use the linkage results. In sum, test developers should be FRANK when conducting and reporting the results of a linkage, to better facilitate fair test score use.

If this chapter has made anything clear, it is that defining what is and what is not a fair use of linked test scores is likely to remain a somewhat arbitrary question that is specific to the test(s) at hand. The test characteristics, test administration/delivery conditions, testing population(s), examinee characteristics, linkage conditions, linkage quality, how the linked scores are used, and value judgments of test users will all play a role in determining fairness. As such, the answer to the questions of whether it is fair to compare scores that have been linked across different modes of administration and whether it is fair to compare scores that have been linked across different tests can probably be answered no more definitively than "it depends." What can be stated definitively, however, is that future fairness discussions and evaluations should be broadened to include the types of test score usage addressed here.

Notes

1. The views expressed are those of the author and not necessarily those of the Department of Defense or the United States Government.
2. The *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014) address broader types of linked scores in a chapter on scales, scores, norms, cut scores, and score linking, but not from a context of fairness. The chapter on fairness in testing raises the issue of score comparability pertaining to test accommodations, adaptations, and modifications, but makes no mention of the fairness of comparing test scores across different tests or modes of administration.
3. One exception is Willingham and Cole (1997), who addressed computer-based testing from a fairness perspective in their seminal book on gender fairness.
4. One exception is Dorans (2004a, 2012), who raised some specific fairness concerns associated with linkages between ACT and SAT scores.
5. Predictive equivalence is called for here instead of correlational equivalence, consistent with Dorans' (2004b) notion that differential prediction studies are preferential to differential validity (i.e. differential test/criterion correlation) studies.
6. Holland and Dorans (2006) stated a preference for the term "comparable scales" over "comparable scores" to make it clear that it is the score distributions that have been made comparable, not the scores (N. Dorans, personal communication, October 17, 2014). Because the focus here is on scores and score usage, the term comparable scores will be used rather than comparable scales.
7. The meaning assigned here to the term calibration is not to be confused with other meanings that have historically been applied, including vertical scaling and estimating item response theory item parameters to be on a common scale (Holland, 2007; Kolen & Brennan, 2014). In the latter case, a variation such as *scalibration* might be more appropriate in the context of linking, since a scaling component is often required to ensure that the calibrated parameters are on the desired scale.
8. Realistic in that institutions are not inclined to validate and use ACT and SAT scores separately, even when advised to do so if feasible (R. Sawyer, personal communication, August 4, 2014).
9. Concerns about the classification of individuals into groups (addressed earlier) are relevant here.
10. An SAS macro to compute systemized score equity assessment is presented in Moses, Liu, and Dorans (2010).
11 Linkages between ACT and SAT scores were conducted in 2010, 1997, 1991, and earlier. The current ACT-SAT concordances will need to be updated once again, following substantial changes to the SAT projected for 2016. The need for updated concordances highlights another problem with linkages between different tests; namely, every time a content or scoring change is made to one of the tests, the linkage needs to be updated. The upside, however, is that the instability of linkages over time is less likely to be a concern the more frequently a linkage is updated.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

American Psychological Association (1986). *Guidelines for computer-based tests and interpretations*. Washington, DC: American Psychological Association.

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 508–600). Washington, DC: American Council on Education.

Bartram, D. (2006). Testing on the Internet: Issues, challenges and opportunities in the field of occupational assessment. In D. Bartram & R. K. Hambleton (Eds.), *Computer-based testing and the Internet: Issues and advances* (pp. 13–37). Hoboken, NJ: John Wiley & Sons.

Baumer, M., Roded, K., & Gafni, N. (2009). Assessing the equivalence of internet-based vs. paper-and-pencil psychometric tests. In D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC conference on computerized adaptive testing*. Retrieved from www.psych.umn.edu/psylabs/CATCentral

Bennett, R.E. (2003). *Online assessment and the comparability of score meaning* (Research Memorandum RM-03-05). Princeton, NJ: Educational Testing Service.

Bergstrom, B. A. (1992, April). *Ability measure equivalence of computer adaptive and pencil and paper tests: A research synthesis*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Blazer, C. (2010). *Computer-based assessments* (Information Capsule Vol. 0918). Miami, FL: Miami-Dade County Public Schools.

Breland, H. M., & Ironson, G. H. (1976). Defunis reconsidered: A comparative analysis of alternative admissions strategies. *Journal of Educational Measurement, 13*, 89–99.

Brennan, R. L. (2007). Tests in transition: Discussion and synthesis. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 161–169). New York: Springer-Verlag.

Bridgeman, B., Lennon, M. L., & Jackenthal, A. (2003). Effects of screen size, screen resolution, and display rate on computer-based test performance. *Applied Measurement in Education, 16*, 191–205.

Bugbee, A. C. (1996). The equivalence of paper-and-pencil and computer-based testing. *Journal of Research on Computing in Education, 28*, 282–289.

Camilli, G. (1993). The case against item bias detection techniques based on external criteria: Do item bias procedures obscure test fairness issues? In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 397–417). Hillsdale, NJ: Lawrence Erlbaum.

Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 221–256). Westport, CT: American Council on Education/Praeger.

Choi, S. W., & Tinkler, T. (2002, April). *Evaluating comparability of paper-and-pencil and computer-based assessment in a K-12 setting*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Chua, Y. P. (2012). Replacing paper-based testing with computer-based testing in assessment: Are we doing wrong? *Procedia: Social and Behavioral Sciences, 64*, 655–664.

Cizek, G. J. (1999). *Cheating on tests: How to do it, detect it, and prevent it*. Mahwah, NJ: Lawrence Erlbaum.

Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and White students in integrated colleges. *Journal of Educational Measurement, 5*, 115–124.

Cole, N. S., & Zieky, M. J. (2001). The new faces of fairness. *Journal of Educational Measurement, 38*, 369–382.

Cronbach, L. J. (1980). Validity on parole: How can we go straight? In W. B.Schrader, (Ed.), *New directions for testing and measurement: Measuring achievement: Progress over a decade* (No. 5, pp. 99–108). San Francisco: Jossey-Bass.

Darlington, R. B. (1971). Another look at "cultural fairness." *Journal of Educational Measurement, 8*, 71–82.

Dorans, N. J. (2004a). Equating, concordance, and expectation. *Applied Psychological Measurement, 28*, 227–246.

Dorans, N. J. (2004b). Using population invariance to assess test score equity. *Journal of Educational Measurement, 41*, 43–68.

Dorans, N. J. (Ed.) (2004c). Assessing the population sensitivity of equating functions. [Special Issue]. *Journal of Educational Measurement, 41*(1).

Dorans, N. J. (2012). The contestant perspective on taking tests: Emanations from the statue within. *Educational Measurement: Issues and Practice, 31*(4), 20–37.

Dorans, N. J. (2013). On attempting to do what Lord said was impossible: Commentary on van der Linden's "Some conceptual issues in observed-score equating." *Journal of Educational Measurement, 50*, 304–314.

Dorans, N. J., & Holland, P. W. (2000). Population invariance and the equitability of tests: Basic theory and the linear case. *Journal of Educational Measurement, 37*, 281–306.

Dorans, N. J., & Liu, J. (2009). *Score equity assessment: Development of a prototype analysis using SAT mathematics test data across several administrations* (ETS Research Rep. No. RR-09-08). Princeton, NJ: Educational Testing Service.

Dorans, N. J., & Middleton, K. (2012). Addressing the extreme assumptions of presumed linkings. *Journal of Educational Measurement, 49*, 1–18.

Dorans, N. J., & Petersen, N. S. (2010, April–May). *Distinguishing concordances from equating.* Paper presented at the annual meeting of the National Council on Measurement in Education, Denver, CO.

Dorans, N. J., Pommerich, M., & Holland, P. W. (Eds.). (2007). *Linking and aligning scores and scales.* New York: Springer-Verlag.

Dorans, N. J., & Walker, M. E. (2007). Sizing up linkages. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 179–198). New York: Springer-Verlag.

Dorans. N. J., Lyu, C. F., Pommerich, M., & Houston, W. M. (1997). Concordance between ACT assessment and recentered SAT I sum scores. *College and University, 73*(2), 24–34.

Downing, S. M., & Haladyna, T. M. (1996). A model for evaluating high-stakes testing programs: Why the fox should not guard the chicken coop. *Educational Measurement: Issues and Practice, 15*, 5–12.

Drasgow, F., & Chuah, S. C. (2005). Computer-based testing. In E. Diener & M. Eid (Eds.), *Multimethod measurement in psychology* (pp. 87–100). Washington, DC: American Psychological Association.

Drasgow, F., Nye, C. D., Guo, J., & Tay, L. (2009). Cheating on proctored tests: The other side of the unproctored debate. *Industrial and Organizational Psychology, 2*, 46–48.

Eignor, D. R. (2007). Linking scores derived under different modes of test administration. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 135–159). New York: Springer.

Feuer, M. J., Holland, P. W., Green, B. F., Bertenthal, M. W., & Hemphill, F. C. (1999). *Uncommon measures: Equivalence and linkage among educational tests.* Washington, DC: National Academy Press.

Flanagan, J. C. (1951). Units, scores, and norms. In E. F. Lindquist (Ed.), *Educational measurement* (1st ed., pp. 695–763). Washington, DC: National Academy Press.

Green. B. F. (1984). Construct validity of computer-based tests. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 77–86). Hillsdale, NJ: Lawrence Erlbaum.

Hanson, B. A., Harris, D. J., Pommerich, M., Sconing, J. A., & Yi, Q. (2001). *Suggestions for the evaluation and use of concordance results* (ACT Research Report No. 2001-1). Iowa City, IA: ACT.

Hartigan, J. A., & Wigdor, A. K. (Eds.) (1989). *Fairness in employment testing: Validity generalization, minority issues, and the General Aptitude Battery.* Washington, DC: National Academy Press.

Holland, P. W. (2007). A framework and history for score linking. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 5–30). New York: Springer.

Holland, P. W., & Dorans, N. J. (2006). Linking and equating. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 187–220). Westport, CT: American Council on Education/Praeger.

Huff, K. L., & Sireci, S. G. (2001). Validity issues in computer-based testing. *Educational Measurement: Issues and Practice, 20*(3), 16–25.

Huggins, A. C., & Penfield, R. D. (2012). An NCME instructional module on population invariance in linking and equating. *Educational Measurement: Issues and Practice, 31*(1), 27–40.

Joint Committee on Testing Practices (2004). *Code of fair testing practices in education (revised).* Washington, DC: Joint Committee on Testing Practices.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*, 1–73.

Kapoor, S., & Welch, C. (2011, April). *Comparability of paper and computer administrations in terms of proficiency interpretations.* Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Karkee, T., Kim, D. I., & Fatica, K. (2010, April–May). *Comparability study of online and paper and pencil tests using modified internally and externally matched criteria.* Paper presented at the annual meeting of the American Educational Research Association, Denver, CO.

Kolen, M. J. (1999). Threats to score comparability with applications to performance assessments and computerized adaptive tests. *Educational Assessment, 6*, 73–96.

Kolen, M. J. (2004). Linking assessments: Concept and history. *Applied Psychological Measurement, 28*, 219–226.

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). New York: Springer.

Kunnan, A. J. (2000). Fairness and justice for all. In A. J. Kunnan (Ed.), *Fairness and validation in language assessment* (pp. 1–14). Cambridge: Cambridge University Press.

Langenfeld, T. E. (1997). Test fairness: Internal and external investigations of gender bias in mathematics testing. *Educational Measurement: Issues and Practices, 16*(1), 20–26.

Lindquist, E. F. (1964, February). *Equating scores on non-parallel tests.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Linn, R. L. (1984). Selection bias: Multiple meanings. *Journal of Educational Measurement, 21*, 33–47.

Linn, R. L. (1993). Linking results of distinct assessments. *Applied Measurement in Education, 6*, 83–102.

Liu, J., & Dorans, N. J. (2013). Assessing a critical aspect of content continuity when test specifications change or test forms deviate from specifications. *Educational Measurement: Issues and Practice, 32*(1), 15–22.

Liu, J., & Walker, M. E. (2007). Score linking issues related to test content changes. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 109–134). New York: Springer-Verlag.

Liu, J., Cahn, M. F., & Dorans, N. J. (2006). An application of score equity assessment: Invariance of linkage of new SAT to old SAT across gender groups. *Journal of Educational Measurement, 43*, 113–129.

Liu, J., Dorans, N. J., & Moses, T. (2010, April–May). *Evaluating the subpopulation sensitivity of the ACT-SAT concordances.* Paper presented at the annual meeting of the National Council on Measurement in Education, Denver, CO.

Lottridge, S., Nicewander, A., Schulz, M., & Mitzel, H. (2008). *Comparability of paper-based and computer-based tests: A Review of the methodology.* Monterey, CA: Pacific Metrics Corporation.

Mazzeo, J., Druesne, B., Raffeld, P. C., Checketts, K. T., & Muhlstein, A. (1991). *Comparability of computer and paper-and-pencil scores for two CLEP general examinations* (College Board Rep. No. 91-5). New York: College Entrance Examination Board.

Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin, 9*, 287–304.

Mislevy, R. J. (1992). *Linking educational assessments: Concepts, issues, methods, and prospects.* Princeton, NJ: ETS Policy Information Center.

Moses, T., Liu, J., & Dorans, N. (2010). Systemized SEA in SAS. *Applied Psychological Measurement, 24*, 552–553.

Mroch, A. A., Li, D., & Thompson, T. D. (2015, April). *A framework for evaluating score comparability*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.

Naglieri, J. A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A., & Velasquez, R. (2004). Psychological testing on the Internet: New problems, old issues. *American Psychologist, 59*, 150–162.

NCME Ad Hoc Committee on the Development of a Code of Ethics. (1995). *Code of professional responsibilities in educational measurement*. Washington, DC: National Council on Measurement in Education.

Neuman, G., & Baydoun, R. B. (1998). Computerization of paper-and-pencil tests: When are they equivalent? *Applied Psychological Measurement, 22*, 71–83.

Paek, P. (2005). *Recent trends in comparability studies* (PEM Research Report No. 05-05). Iowa City, IA: Pearson Educational Measurement.

Petersen, N. S., & Novick, M. R. (1976). An evaluation of some models for culture-fair selection. *Journal of Educational Measurement, 13*, 3–29.

Pommerich, M. (2004). Developing computerized versions of paper-and-pencil tests: Mode effects for passage-based tests. *Journal of Technology, Learning, and Assessment, 2*(6). Retrieved from www.jtla.org

Pommerich, M. (2007a). The effect of using item parameters calibrated from paper administrations in computer adaptive test administrations. *Journal of Technology, Learning, and Assessment, 5*(7). Retrieved from www.jtla.org

Pommerich, M. (2007b). Concordance: The good, the bad, and the ugly. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 199–216). New York: Springer-Verlag.

Pommerich, M. (2012). Comments on Neil Dorans's NCME career award address: The contestant perspective on taking tests: Emanations from the statue within. *Educational Measurement: Issues and Practice, 31*(4), 40–44.

Pommerich, M., & Dorans, N. J. (Eds.) (2004). Concordance. [Special Issue]. *Applied Psychological Measurement, 28*(4).

Pommerich, M., Hanson, B. A., Harris, D. J., & Sconing, J. A. (2004). Issues in conducting linkages between distinct tests. *Applied Psychological Measurement, 28*, 247–273.

Randall, J., Sireci, S., Li, X., & Kaira, L. (2012). Evaluating the comparability of paper- and computer-based science test across sex and SES subgroups. *Educational Measurement: Issues and Practice, 31*(4), 2–12.

Rudner, L. M. (2010). Implementing the Graduate Management Admission Test computerized adaptive test. In W. J. van der Linden (Ed.), *Elements of adaptive testing* (pp. 151–165). New York: Springer.

Sands, W. A., Waters, B. K., & McBride, J. R. (Eds.) (1997). *Computerized adaptive testing: From inquiry to operation*. Washington, DC: American Psychological Association.

Sawaki, Y. (2001). Comparability of conventional and computerized tests of reading in a second language. *Language Learning & Technology, 5*, 38–59.

Sawyer, R. (2007). Some further thoughts on concordance. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 217–230). New York: Springer-Verlag.

Sawyer, R., Cole, N. S., & Cole, J. W. L. (1976). Utilities and the issue of fairness in a decision theoretic model for selection. *Journal of Educational Measurement, 13*, 59–76.

Schaeffer, G. A., Steffen, M., Golub-Smith, M. L., Mills, C. N., & Durso, R. (1995). *The introduction and comparability of the computer adaptive GRE General Test* (GRE Board Report No. 88-08aP). Princeton, NJ: Educational Testing Service.

Schroeders, U., & Wilhelm, O. (2011). Equivalence of reading and listening comprehension across test media. *Educational and Psychological Measurement, 71*, 849–869.

Stokes, K. (2013, July 29). Review: On average, online disruptions didn't hurt Indiana's ISTEP+ scores. *Indiana Public Media*. Retrieved from http://indianapublicmedia.org

Texas Education Agency (2008). *A review of literature on the comparability of scores obtained from examinees on computer-based and paper-based tests*. Texas Education Agency (TEA) Technical Report Series.

Thissen, D. (2007). Linking assessments based on aggregate reporting: Background and issues. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 287–312). New York: Springer-Verlag.

Thorndike, R. L. (1971). Concepts of cultural fairness. *Journal of Educational Measurement, 8*, 63–70.

Tippins, N. T. (2009). Internet alternatives to traditional proctored testing: Where are we now? *Industrial and Organizational Psychology: Perspectives on Science and Practice, 2*, 2–10.

Tippins, N. T., Beaty, J., Drasgow, F., Gibson, W. M., Pearlman, K., Segall, D. O., & Shepard, W. (2006). Unproctored Internet testing in employment settings. *Personnel Psychology, 59*, 189–225.

U.S. Census Bureau. (2014). *Computer and Internet trends in America*. Retrieved from www.census.gov/hhes/computer/files/2012/Computer_Use_Infographic_FINAL.pdf

von Davier, A. A., & Liu, M. (2008). Population invariance. [Special Issue]. *Applied Psychological Measurement, 32*(1).

Wang, H, & Shin, C. D. (2010). *Comparability of computerized adaptive and paper-pencil tests*. Test, Measurement & Research Services Bulletin, Issue 13. Pearson Education.

Wang, S., Jiao, H., Young, M. J., Brooks, T. E., & Olson, J. (2007). A meta-analysis of testing mode effects in Grade K-12 mathematics tests. *Educational and Psychological Measurement, 67*, 219–238.

Wang, S., Jiao, H., Young, M. J., Brooks, T., & Olson, J. (2008). Comparability of computer-based and paper-and-pencil testing in K-12 reading assessments. *Educational and Psychological Measurement, 68*, 5–24.

Wang, T., & Kolen, M. J. (2001). Evaluating comparability in computerized adaptive testing: Issues, criteria and an example. *Journal of Educational Measurement, 38*, 19–49.

Willingham, W. W., & Cole, N. S. (1997). *Gender and fair assessment*. Mahwah, NJ: Lawrence Erlbaum.

Winter, P. C. (2010). Comparability and test variation. In P. C. Winter (Ed.), *Evaluating the comparability of scores from achievement test variations* (pp. 1–11). Washington, DC: Council of Chief State School Officers.

Wolfe, J. H., Moreno, K. E., & Segall, D. O. (1997). Evaluating the predictive validity of CAT-ASVAB. In W. A. Sands, B. K. Waters, & J. R. McBride (Eds.), *Computerized adaptive testing: From inquiry to operation* (pp. 175–179). Washington, DC: American Psychological Association.

Xi, X. (2010). How do we go about investigating test fairness? *Language Testing, 27*(2), 147–170.

Yin, P., Brennan, R. L., & Kolen, M. J. (2004). Concordance between ACT and ITED scores from different populations. *Applied Psychological Measurement, 28*, 274–289.

Zieky, M. (2006). Fairness reviews in assessment. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 359–376). Mahwah, NJ: Lawrence Erlbaum.

Zitny, P., Halama, P., Jelinek, M., & Kveton, P. (2012). Validity of cognitive ability tests: Comparison of computerized adaptive testing with paper and pencil and computer-based forms of administrations. *Studia Psychologica, 54*, 181–194.

Comments of Deb Merritt for OEP

Thoughts from Debby Merritt

Formatted: Left

OEP Admissions Qualifications/Selection

From the original Supreme Court submission:

> Finally, we recommend expressly encouraging holistic admission practices including admitting law students on more than an evaluation of LSAT/GPA in order to ensure reliance on more inclusive criteria, such as work experience, life experience, and/or overcoming personal challenges. Law schools will inherently be encouraged to do so if they have the confidence that all first-year students can apply for the OEP program. Accordingly, we recommend making clear that the OEP will be open to all students in the spring of 1L year (rather than limiting participation to those pre-selected for the program)

- Do not have general statements about law school admissions
  - Requiring or prescribing admissions processes
  - How might OEP change their regular admissions process

Common Qualifications:

- GPA: Good academic standing vs. number (not uniform)
- Writing skills
- Process for holistic review of candidates
- Foundational skill qualifications (lawyer vs law student) (possibly evidence based)
- Consider difference between a helpful list vs subjective indicators

I support each law school developing its own admission rules for the OEP. I don't think these need to be uniform across the schools, especially since they are likely to have different capacities for the OEP.

Clinics at some law schools use an application process that might work well for the OEP. Common requirements are for students to submit a CV and statement of interest indicating why they are interested in the program (here, the OEP) and how it relates to their future plans.

I recommend *eliminating* any consideration of class rank, GPA, or LSAT. Applicants should omit those references from CVs submitted to the OEP. Grades from first-year writing or lawyering classes might seem particularly relevant, but first-generation students often struggle with these classes. For that reason, I would omit consideration of even those grades.

Nor would I require a writing sample. Law practice involves many tasks and types of communication. A writing sample, especially one completed during the first year, may not be

representative of what a student can achieve in the OEP. And again, first-generation students often take longer than others to learn the rules of legal writing. They shouldn't be disadvantated by that.

I would consider giving preference for the OEP to students who are not on the law review. Law review students already enjoy a premium experience from the school; other students should have a first shot at this different premium experience (the OEP). Law review students are also more likely to want the portability provided by taking the UBE (and to succeed on that exam). Writing for the law review also demands time that might conflict with participation in the OEP. To the extent that law review membership depends on grades, finally, this reverse preference may assure that students who will struggle most with the UBE will have access to the OEP.

A final thought: If there is any discretion in who is admitted (e.g., if personal statements are used), then it's important to decide who is on the admissions committee. Race and gender diversity on that committee will be important, as well as a commitment to DEI among all members (maybe with some training). I would include at least one student currently enrolled at the OEP at that school.

Requirements:

Documentation:

Advising:

Entry: Preliminary program plan

Ongoing: Regular check-ins to measure ongoing success/challenges in program

Meeting requirements to continue program

Outboarding

Others:

Email Exchange between Jen Reynolds and Troy Wood regarding OEP

Troy Wood

From: Jen Reynolds <jwr@uoregon.edu>
Sent: Wednesday, September 14, 2022 4:01 PM
To: Troy Wood
Subject: Re: OEP Working Notes

Thanks Troy. What is the "3L work product"? Is that something separate from their work in the designated courses?

From: Troy Wood <twood@osbar.org>
Date: Wednesday, September 14, 2022 at 2:40 PM
To: Jen Reynolds <jwr@uoregon.edu>
Cc: Rebekah Hanley <rhanley@uoregon.edu>
Subject: RE: OEP Working Notes

Jen:

I am sorry that these answers were not more apparent. It was presumed that many of the institutions would have their original rep on this committee as well. Most, if not all of your questions were answered by the Task Force, or the Court, prior to sending to the Committee. Please see answers below in red, and let us know what input you have following these answers. I am BCC'ing Tony, so he can be aware of some of the ambiguity created by not including the Task Force decisions in our prior discussions. However, I would discourage anyone from hitting REPLY ALL. I am also blind copying Stuart, as he is a Committee Member, and I do not want any voting members to be engaged in a pre-meeting dialogue.

Best to all,

Troy

Oregon State Bar
Troy Wood
Regulatory Counsel
503-431-6310
twood@osbar.org

Oregon State Bar • 16037 SW Upper Boones Ferry Road • PO Box 231935 • Tigard, OR 97281-1935 • www.osbar.org

---

From: Jen Reynolds <jwr@uoregon.edu>
Sent: Wednesday, September 14, 2022 10:02 AM
To: Troy Wood <twood@osbar.org>
Cc: Stuart Chinn <schinn@uoregon.edu>; Rebekah Hanley <rhanley@uoregon.edu>
Subject: Re: OEP Working Notes

Hi Troy! Just a few thoughts after speaking with my colleagues and co-members of the committee, Stuart and Rebekah:

Before we can provide feedback on a document like this one, we need to have more of a sense of the following:

- **Who is eligible to participate in the OEP?** All Oregon law students who complete 1L and meet the qualifications for the program, which is why we are discussing qualifications now, because they will decide the limits of capacity. However, the rules will make the qualifications broad, and it will be up to the students as to what program path they take, and school personnel to steer students into programs that are best for them. **If the answer is "second semester 1L students," then how many can participate? What are the parameters of the OEP program (e.g., number and type of courses) and who does what assessments when?** This is one of the primary purposes of this committee (to determine what the curriculum looks like). There was model presented in the task force memo that included the following:

  - **Foundational Courses Beyond the Required First Year Courses:** (range: ~20-24 credits, noting that credits assigned by the law schools for completing these courses can vary)
    - Successful completion of the following foundational upper-level courses:
      - Professional Responsibility (2-3)
      - Evidence (3-4)
      - Two of the following:
        - state/ local law (2-3),
        - constitutional or statutory interpretation (2-3), or
        - administrative law or processes (2-3).
    - Take 3 of the following:
      - Criminal Procedure (3),
      - Business Associations (3),
      - Family Law (3),
      - Trusts & Estates (3),
      - Personal Income Tax (3).
    - Successful completion of a graduate writing requirement (2-3 credits) that complies with ABA Standard 303(a)(2).
  - **Experiential Requirements** (15 credits)
    - Successful completion of no fewer than 9 credits of closely supervised clinical work or simulation coursework.
    - Successful completion of up to 6 credits of externship work.
  - **EAP Capstone Requirement** (for development by the OEP Implementation Task Force during AY21-23)
    - To be developed in partnership with BBX. We could imagine, for instance, the creation of performance tests using case files and a limited universe of materials. We could alternatively imagine creation of a capstone project that relies on a rubric generated by the OEP Implementation Task Force. The rubric could serve as a curricular planning tool for students and, in doing so, could permit development of EAPs that students could begin during the fall of their second year. That rubric should consider the building blocks of minimum skills competence alongside ABA learning outcomes.
    - The requirements in this curriculum, including the first-year, total between 65-69 credits, although most of those requirements allow considerable choice among subject areas. Since ABA accreditation standards require at least 83 credits of academic work to secure a J.D., the course requirements in this example permit at least 14 credits (i.e., a full semester) of completely elective courses. The system, in other words, structures the JD program while still allowing considerable student choice.

Without an idea of the resources required for the OEP or an OEP pilot, it's hard to know how to determine eligibility/numbers. The BBX will review 3L work product. The rules will have requirements for staying in the program. Faculty will review student work product while in the first year of the program, and it will be up to schools to review and score the applicants material and determine if they meet the requirements as laid out in

the rules. However, the rules will give a great deference to the schools in setting criteria for maintaining this pathway.

- **How does a student apply to the OEP?** It will be a bifurcated application. The BBX will have a C&F application, that the applicant will have to fill out. The school will have their own applications for applicants who meet the qualifications. The school will need to provide a copy of the application to the BBX when they have accepted the applicant into their program. The BBX will be then verify with the school that the applicant has also filled out the C&F part of the application. **Who develops and maintains the application process (website, forms, etc.)? Who provides information to the students about the program, advises them on whether it makes sense for them, etc.? (In fact, how would a student know whether the OEP was a better option than the bar or the SPP?) If interviews are required, who conducts the interviews?** C&F will be exclusively handled by the BBX, and the program application will be maintained by the school, and whether an applicant has been able to remain in the program will be maintained by the school, and letting the BBX know if an applicant moves out of the program. The BBX would recommend that the applicants be required to fill out an application for the third year as well, which will give a second opportunity for the school to help applicants who won't make it to licensure in this pathway to another pathway.

- **How are students selected to participate in the OEP? Who develops and articulates the process and criteria? Whose job is it to review applications, make selections, announce results, and manage rejections/disputes/appeals?** This will be an internal school process. The BBX will not dictate how schools perform this function.

- **Do we envision that everyone in the OEP will successfully complete the program (so long as they don't drop out of the OEP voluntarily), or do we envision a failure rate (as we have with the bar exam)?** In BBX discussions on this subject, it is anticipated that there will be a fail rate of some sort initially. The size of the rate could be almost zero depending on the institution's ability to establish its own cut between the 2L & 3L years, or the school's ability to course correct applicants who will not meet the standard upon graduation. If an applicant's work product does not measure up to the competence standard on graduation, then they will fail at this pathway, but they will be able to obtain licensure through the bar exam or the SPP program.

- **How do we ensure that the OEP provides consistent results across the law schools, given differences in the areas of strength, faculty resources, grade curves, etc.? What common threshold requirements are possible, especially since academic freedom means that we cannot dictate what faculty members teach in their courses? Who will be responsible for developing and measuring these threshold requirements?** The BBX will perform the gatekeeper function as they do now on the bar exam. They will be the ones who determine if the work product meets the professional standard. As the standard will be assessed using methods approved by the Court, as suggested by this committee. I would suspect that the process will be proposed by one of the psychometricians liaisons to this committee. Through that process, the Court will need to be convinced that the BBX will be able to hold this line from school to school and cohort to cohort over many successful years of the OEP program.

- **Will employers/clients know whether their lawyer did the bar exam or the SPP or the OEP?** The pathway used for licensure has always been a public record. I seriously doubt this will change with these new programs. However, if New Hampshire offers any forecast as to what we can expect in Oregon, then the OEP program will probably become the sought after method to licensure.

Thanks very much!
Jen Reynolds

From: Troy Wood <twood@osbar.org>
Date: Tuesday, September 13, 2022 at 5:26 PM

To: Troy Wood <twood@osbar.org>
Subject: RE: OEP Working Notes

Please remember that tomorrow is the deadline for comments on the OEP outline and framework thus far, so that we can move this process along at our next meeting on 9/21/2022. I have received some meaningful feedback, which we will share at the next meeting, but only a few have responded thus far. All feedback will be helpful.

Thank you,

Troy

Oregon State Bar
Troy Wood
Regulatory Counsel
503-431-6310
twood@osbar.org

Oregon State Bar • 16037 SW Upper Boones Ferry Road • PO Box 231935 • Tigard, OR 97281-1935 • www.osbar.org

From: Troy Wood
Sent: Wednesday, September 07, 2022 9:40 PM
To: Troy Wood <twood@osbar.org>
Subject: OEP Working Notes

Dear OEP Workgroup and Interested Parties:

You are receiving these working notes from the OEP program lead, Dr. Anthony Rosilez, because the OEP group is trying to develop its work product by seeking advice from individuals who may have a special background in areas where it needs some input; or from its SPP counterparts, who may have addressed these issues in its work.

Please see attached working notes, to provide any input that you feel may help the OEP group in its discussions. If you enter any notes, please provide your thought process for entering the notes, and or provide any supporting articles or authorities that support your note.

Please send your response to me alone, and do not copy any other participants, as this might be seen as an effort to hold an electronic public meeting, which we would like to avoid.

I will need your responses within a week, as this will allow Tony and I to add a final work product to the agenda for the 9/21/2022 public meeting of the LPDC.

Thank you in advance for any thoughts you might share.

Troy

Oregon State Bar

Troy Wood
Regulatory Counsel
503-431-6310
twood@osbar.org

Oregon State Bar • 16037 SW Upper Boones Ferry Road • PO Box 231935 • Tigard, OR 97281-1935 • www.osbar.org