inflexion

Exploring the Practices for and Effects of Open Book Assessment

Final Report

June 2021

inflexion
360 E 10th Ave Suite 300
Eugene, OR 97401
1.541.246.2600

Prepared by
Jessica N. Jacovidis, PhD
Kristine L. Chadwick, PhD
Paul Beach, PhD

Table of Contents

List of Tables

List of Figures

Executive Summary

In July 2020, the IB contracted with Inflexion to explore the effects and practices for open book examinations (OBEs). This study sought to inform IB's understanding of the best practices in the use of resource aids from a practical perspective (e.g., students with specific educational and assessment needs, cultural differences, academic integrity, teacher assessment literacy and expertise, and school resources). Ultimately, the goal of the study was to develop guidelines specific to OBEs and use of reference aids on IB summative and classroom-based assessments anchored in the best available research evidence and contextualized to the IB. This study addressed the following five overarching research questions; some of the research questions had additional sub-questions:

1. What are the effects of the use of reference aids in high-stakes summative assessment, compared to assessment without these aids?
2. What are the effects of using reference aids in classroom assessment, compared to assessment without these aids?
3. What are good practices regarding assessments using reference aids for high-stakes summative assessment?
4. What are good practices regarding classroom assessment and examination preparation for assessments using reference aids and how can these be made inclusive for all students?
5. What are the essential teaching and assessment design skills, and school resources required to implement assessment using reference aids in the classroom?

Research Design and Methodology

Inflexion researchers employed a sequential, multiphase mixed methods design. In Phase 1, Inflexion conducted a comprehensive literature review by exploring academic databases and the websites of targeted organizations. In Phase 2, Inflexion performed a targeted review of IB documents, conducted interviews with key IB staff, and administered a survey of IB stakeholders. Taken together, the data from Phase 1 and Phase 2 were used to develop guidelines on OBEs and the use of reference aids within IB.

Phase 1 Findings

- **Limited Literature:** The literature base for OBEs is unsystematic, uncomprehensive, and splintered. Most of the research on OBEs is correlational, based on small-scale interventions in higher education, and lacks sufficient detail about the nature of the reference aids examined, with many studies simply noting students had access to reference materials.
- **Student Performance:** Results on student performance are mixed and highly dependent on context. OBEs and closed book exams (CBEs) are generally compared in terms of long-term retention; however, this type of comparison may not be appropriate, as it does not align with the purpose of OBEs.
- **Student Test Preparation:** Research is inconclusive on whether students prepare differently for OBEs and CBEs. In the absence of clear expectations or specific training, students will likely prepare less for OBEs than CBEs. Preparation may also vary by students' learning orientation.
- **Student Well-being:** Research on student anxiety and stress relies almost exclusively on student self-report responses to surveys or interviews. Generally, students report less anxiety and reduced stress from OBEs relative to CBEs; however, OBEs can increase anxiety and stress when students underestimate the difficulty of OBEs, are unfamiliar with the OBE format, or have little experience preparing and using reference aids.
- **Backwash Effects on Learning:** OBEs can provide powerful backwash effects on learning. For this to be true, a number of factors must be addressed, including designing assessments to target higher order thinking skills, providing clear expectations around the use of reference aids, and providing teachers and students training and practice with OBEs.
- **Assessment Objectives/Purpose:** Well-designed OBE questions require students to go beyond factual knowledge to demonstrate higher order thinking, problem solving, application, and analysis. Teaching, learning, and assessment format and item design must all be aligned to effectively target and measure higher order thinking on OBEs.
- **Exam and Item Design:** OBEs can be designed and administered in ways that do not sacrifice psychometric quality. However, particular attention should be given to balancing breadth and depth of content coverage, reliability, and the impact on test score validity.
- **Assessment literacy and design competency:** Very little research exists on the type of assessment design skills needed to effectively implement OBEs.
- **Academic Integrity:** Clear expectations, honor code agreements, and cheating detection software can be used to combat academic dishonesty on OBEs.
- **School resources:** Very little research exists on what type of school resources are needed to support and train teachers to effectively implement OBEs.

Phase 2 Findings

- **Receptivity:** Generally, receptiveness to OBEs and the use of reference aids is high. However, the purpose of the assessment is key. IB staff and stakeholders recognize that reference aids should only be used if it is the best way to assess the knowledge and skills that are being targeted by the assessment.

- **Alignment with IB Philosophies:** The use of reference aids aligns with existing IB policies and pedagogical practices. Further, IB is already using reference aids in some content areas and has some existing policies around specific reference aids.
- **Ease of Incorporation of OBEs:** Reference aids should be fairly easy to incorporate into IB's practice. However, there are too many unknowns at this point to fully understand what the implementation of OBEs would look like for IB.
- **Suitability of Reference Aids for Different Subjects:** Some reference aids may be better suited for certain subject areas. However, it is difficult to know *which* reference aids will be best suited for what subject areas.
- **Perceived Impact of OBEs:** IB staff and stakeholders generally felt introducing OBEs would have positive effects, including increasing assessment authenticity and the ability to assess higher order thinking instead of recall and memorization, and potentially, decreasing student stress.
- **Relevance of Recommendations from the Literature:** Even with the limited literature base, the recommendations from research seem to be relevant for IB's work.
- **Concerns Regarding OBEs:** There are some concerns about the effects of OBEs on academic honesty and equity. However, most other concerns from interviewees stemmed from a lack of clarity around what reference aids may be used and in what subject areas.
- **Assessment Skills for OBEs:** Assessment-related skills will be key for the success of OBEs. However, there is limited information on the teaching and assessment design skills that are necessary for teachers to use reference aids effectively.
- **Professional Learning:** Professional learning will be needed both internally and externally. There are too many unknowns regarding how IB might proceed with OBEs and the use of reference aids to know how this might affect schools and IB's work. Training sessions, templates, and examples will be invaluable.
- **Impact on IB Work:** Given there are still several decisions to be made, IB staff were unaware of how OBEs would impact their work.

Guideline Development and Working Session Conclusions

Inflexion researchers developed initial guidelines specific to OBEs and use of reference aids on IB summative and classroom-based assessments and held three working sessions with IB staff. The goal of the working sessions was to react to the initial guidelines, further articulate the suggested guidelines, and add relevant guidelines or identify areas of work for which guidelines still need to be developed. The development of actionable guidelines was dependent on the quality of information available in the literature, and on specific decisions points for IB. Thus, the guidelines were not as specific and actionable as originally envisioned. The following guidelines present higher-level recommendations and considerations that will need to be further articulated as IB's understanding of OBEs and decisions around their use evolve. Further, Inflexion researchers provided a subjective categorization of each guideline based on the level of importance for successful implementation of OBEs. This coding was designed to facilitate the prioritization of the guidelines and assist IB staff in deciding where to focus their resources.

Importance	Guideline
Overarching and Policy-Related Guidelines	
Critical	1. Provide clear rules and guidelines and set clear expectations for OBEs and the use of reference aids.
Critical	2. Provide professional learning opportunities to all stakeholders.
Critical	3. Review and update policies around accessibility to cover the use of reference aids.
Curriculum Development Guidelines	
Critical	4. Allocate time to teach the skill of creating and/or using reference aids.
Moderate	5. Determine whether use of reference aids should be a skill targeted by the assessment.
Moderate	6. Provide guidance for teachers on preparing students for OBEs.
Assessment Design Guidelines	
Moderate	7. Use reference aids when they are the best choice for assessing the knowledge and skills of interest.
Critical	8. Leverage reference aids to assess higher order thinking skills.
Moderate	9. Ensure that time constraints are realistic given the number of questions and number of reference aids.
Assessment Implementation Guidelines	
Moderate	10. Provide explicit instructions for checking reference aids and exam invigilation.
Moderate	11. Review and update scoring models as needed to account for OBEs and the use of reference aids.
Critical	12. Ensure exams allowing reference aids still meet adequate assessment quality standards.
Professional Learning Guidelines	
Critical	13. Provide learning opportunities on how to write assessment questions that optimize the use of reference aids.
Moderate	14. Provide learning opportunities on how to create and/or use reference aids.
Moderate	15. Provide learning opportunities on checking reference aids and exam invigilation.

Areas for Future Research

Given the lack of information in the literature, there are a number of future research studies that could be conducted to better understand how OBEs and the use of reference aids would work in IB's context.

- If the IB is interested in incorporating OBEs, additional empirical evidence is needed directly comparing key outcomes of interest for open and closed book exams. This information would not only benefit the IB but would contribute significantly to the literature base on open book assessments.
- IB should consider engaging IB teachers as subject matter experts to explore the appropriateness of specific reference aids in content areas for which there is interest in implementing OBEs, and to suggest which existing assessments might benefit from an OBE approach. Once information is gathered from subject matter experts, theoretical findings can be tested empirically. Further, IB could partner with pilot schools to implement plan-do-study-act (PDSA) cycles to sufficiently trial the use of reference aids prior to scaling up.
- IB should consider a study with schools to examine teacher assessment design skills; access to, familiarity with, and use of reference aids; and school resources needed for successful implementation of OBEs. This study could include the guidance that schools and teachers would like to have regarding implementing OBEs in the classroom, and the format in which schools would prefer to receive the guidance and training.
- IB should consider an environmental scan to identify and learn from organizations that have already been successful at implementing OBEs and determine the best way to learn from their experiences. Further, though exemplars exist, IB will need to consider how to tailor these examples to its context.
- IB will need explore the impact of OBEs on its specific tests and in its specific contexts. IB may consider the process that is used to validate exams and/or conduct a series of pilot studies to explore the impact before scaling up the use of OBEs. IB may also want to consider, a priori, an acceptable psychometric benchmark that OBEs must meet.
- IB staff should consider what they would hope to learn about OBEs in paper-pencil and digital assessment, access and inclusion, and testing accommodations and consider how to best capture that knowledge.

Conclusion

Ultimately, this study sought to develop actionable guidelines that could be used as IB expanded OBEs and the use of reference aids within the organization. However, after extensive searching, it became clear that this was not possible. We are coming away with many more questions than answers. There is much more to learn about OBEs before IB can implement OBEs on a large scale. As such, general guidelines and considerations were provided to help IB staff frame their thinking about OBEs and to inform the next steps of the OBE journey.

Exploring the Practices for and Effects of Open Book Assessments

In July 2020, the International Baccalaureate (IB) Organisation contracted Inflexion to explore the effects and practices for open book assessment. The purpose of this study was to develop guidelines specific to Open book examinations (OBEs) and use of reference aids on IB summative and classroom-based assessments anchored in the best available research evidence and contextualized specifically to the IB. This report presents the summative findings of the study and covers work completed between July 2020 and April 2021.

Open Book Examinations (OBEs)

OBEs are not a new concept in education. Writing in *the Journal of Higher Education* in 1934, John and Ruth Stalnaker argued that a typical college exam "demands primarily the recognition or recall of simple isolated factors which, by judicious cramming, can be learned in a short time and forgotten in even less time" (p. 117). The authors favored the OBE format because it "stresses the importance of knowing where to find information, how to evaluate it, and how to use it in problems" (p. 120). In the end, the authors made the following claim:

> Open book examinations will unquestionably be more widely used in the near future. Research is of course necessary to determine what they measure and the form of question best suited to them, and to solve many other problems. Even before experimental work has established the value of these examinations, the study of the logic underlying them suffices to convince one that the open book examination is sound (p. 120).

This statement applies to the present day, nearly 80 years later. Indeed, as John and Ruth Stalnaker predicted in 1934, it appears the value and underlying logic of OBEs has led to the increased presence of OBEs in education. If research activity is any indicator, it appears OBEs have been used more widely since 1934, especially in higher education. And it seems reasonable to suggest OBEs will be used more frequently in the future and at education levels prior to university. For example, a search in ProQuest Education Collection shows the number of articles referencing "open book exam" has more than doubled every decade since the 1980s. However, the literature base surrounding OBEs is fractured and unsystematic. Very little rigorous experimental research has been conducted on OBEs during the last 87 years. That does not mean little research has been conduct on OBEs. In fact, this literature review draws on nearly 200 sources, largely from peer-reviewed academic journals. Despite the limits of unsystematic, mostly correlational, descriptive, and survey-based research, this collection of literature helps produce a picture, albeit incomplete, of OBEs and the contexts in which they are and are not successful.

Beyond research, an increasing number of national governments and international organizations have moved to prioritize higher order thinking skills in curriculum and assessment (Care, Kim, Vista, & Anderson, 2018), which can be more readily assessed using OBEs. The COVID-19 pandemic has forced many organizations that typically offer closed book, paper-and-pencil exams to move their exams to

an online, open book format (e.g., College Board, 2020). Other organizations, such as the IB, are conducting research to better understand if and how OBEs can be leveraged to achieve programmatic goals.

Open Book Examinations for IB Programmes

Founded in 1968, the International Baccalaureate (IB) Organisation is a nonprofit education foundation that offers international education programs to assist students in developing skills needed to succeed in a global world (IB, 2021a). The IB's mission statement sets the organization's intention: "Develop inquiring, knowledgeable and caring young people who help to create a better and more peaceful world through intercultural understanding and respect" by working "with schools, governments and international organizations to develop challenging programs of international education and rigorous assessment" (IB, 2021b, par. 5). As of February 2021, the IB is offered in nearly 5,400 schools in 158 countries (IB, 2021c).

As an international leader in the field of education, IB's "approaches to learning [are] grounded in the belief that learning how to learn is fundamental to a student's education" (IB, 2017, p. 7). Lifelong learners are motivated students who believe they can continually improve their ability to learn using higher order thinking skills. IB programmes are designed to target a set of learner attributes (e.g., inquirers, knowledgeable, thinkers) that aim to "develop internationally minded people who, recognizing their common humanity and shared guardianship of the planet, help to create a better and more peaceful world" (IBO, 2013, p. 1). As IB notes (2019), "Not all aspects of the learner profile are appropriate to measure through summative assessment, but several are encapsulated within the concept of higher-order thinking skills" (p. 71).

OBEs have the potential to expand the number and type of higher order thinking skills targeted in IB exams, thus improving overall construct relevance. However, the potential of OBE to enhance higher order thinking and improve students' well-being is contingent on a host of factors that might have implications for IB schools, as well as internal IB processes and practices. These factors *include but are not limited to* the following:

a. Managing the perceptions of schools and students about OBEs and the use of reference aids
b. Communicating with schools about the expectations of OBEs
c. Providing training to IB staff related to OBEs and the use of reference aids
d. Providing learning opportunities to teachers related to OBEs and the use of reference aids
e. Providing guidance on opportunities for students to practice with various reference aids
f. The impact of OBEs and the use of reference aids on academic integrity and test security
g. The impact of OBEs and the use of reference aids on the DP review and inconsistences across subjects due to different curriculum review schedules
h. The impact of OBEs and the use of reference aids on assessment design and the psychometric properties of the assessment.

Attending to these factors, in tandem, unlocks an additional benefit of transitioning to OBEs: improved overall system coherency.

The Current Study

The current study focused on the trends, practices, and effects of using a variety of reference aids in OBEs. The study had two primary goals. First, this study sought to inform IB's understanding of how the use of reference aids in assessment situations affects student performance and well-being, both in classroom and high-stakes summative assessments. Second, the study was designed to inform IB on the best practices in the use of resource aids from a practical perspective (e.g., students with specific educational and assessment needs, cultural differences, academic integrity, teacher assessment literacy and expertise, and school resources), considering the diverse context in which IB operates. Specifically, this study addressed the following research questions:

1. What are the effects of the use of reference aids in high-stakes summative assessment, compared to assessment without these aids on
    a. student performance and well-being, including that of students with specific assessment needs?
    b. exam and item design, including the discriminatory power of summative assessments?
    c. shifts in assessment objectives, including assessing higher order thinking, practical or ancillary skills assessed using OBE?
    d. academic integrity and expectations regarding referencing?
2. What are the effects of using reference aids in classroom assessment, compared to assessment without these aids on
    a. student performance and well-being, including that of students requiring access arrangements?
    b. the teaching activities and classroom assessment practices, including shifts towards assessing higher order thinking and changes in ancillary skills?
    c. essential teacher assessment literacy and assessment design skills?
    d. school resources?
3. What are good practices regarding assessments using reference aids for high-stakes summative assessment for
    a. assessment model design (i.e., the suite of tests and assessments)?
    b. inclusive exam and item design?
    c. paper-based or technologically enhanced assessment (i.e., digital, onscreen or online)?
    d. assessing of specific disciplines?
    e. monitoring academic integrity?
4. What are good practices regarding classroom assessment and examination preparation for assessments using reference aids (regarding foci in question 3 as appropriate) and how can these be made inclusive for all students?
5. What are the essential teaching and assessment design skills, and school resources required to implement assessment using reference aids in the classroom?

Research Design and Methodology

To address the research questions, Inflexion researchers employed a sequential, multiphase mixed methods design with two distinct phases (see Figure 1). In Phase 1, Inflexion conducted a comprehensive literature review by exploring academic databases and the websites of targeted organizations. In Phase 2, Inflexion performed a targeted review of IB documents, conducted interviews with key IB staff, and administered a survey of IB stakeholders. These data helped to contextualize Phase 1 data, ensuring that the resulting information was highly relevant to IB's international education context, and the recommendations and guidelines were appropriate and relevant to IB policies and practices. Taken together, the data from Phase 1 and Phase 2 were used to develop reference aid guidelines related to curriculum development, assessment design, assessment implementation, and professional learning for various stakeholder groups (e.g., teachers, exam content developers, exam proctors, markers and graders, and IB staff).

Figure 1. Visual representation of the research design.

Phase 1: Comprehensive Literature Review

Phase 1 of this study addressed the first two research questions and consisted of two primary components: an academic database search and a targeted organization search. Each of these components are described in detail below.

Academic Database Search

Inflexion researchers conducted a mixed methods literature review to capture the current state of OBEs and use of reference aids (Grant & Booth, 2009). To guide the literature review, Inflexion researchers developed a working definition of OBEs:

> **Open book assessments are examinations where test takers are permitted to use certain approved resources during the full or part of the testing session. These resources typically include textbooks or student-created notes (e.g., lecture**

notes, crib or "cheat" sheets), but could also include full or limited access to internet sources.

Employing this definition, search terms related to OBEs and specific reference aids were identified and parameters for searches were developed. Inflexion researchers included alternative terms for the resource aids, as needed, based on our knowledge and experience, as well as preliminary research on the identified resource aids. To ensure the most comprehensive set of search terms, Inflexion also worked with IB to determine what terms resonate across cultures and what additional terms might be needed to capture the most globally representative pool of literature.[1] This process resulted in 95 search terms to be explored (see Appendix A for the complete list of search terms).

Inflexion researchers conducted broad-based searches in two academic databases: PsychNet and ProQuest's Education Collection. To focus on the most recent and relevant literature, Inflexion researchers limited the search to article titles and abstracts from 2000–2020. The original 95 search terms resulted in 26,289 potential articles. To maintain a manageable number of articles, Inflexion instituted a two-step search process. We started the academic database search with five broadly defined primary search terms. We then followed up with a secondary set of fifteen search terms that focused on specific reference aids (see Table 1). A precursory review of articles from the remaining search terms indicated that these documents were redundant and/or unrelated to OBEs or use of reference aids. Thus, explicit searches were not performed using these search terms.

Table 1. Search Terms Included in the Academic Database Search

Primary Search Terms
ab("open book" OR "closed book")
ab("test* aid*")
ab ("exam* aid*")
ab("assessment* aid*")
ab(web-based OR internet-based OR paper-based) AND ab(aid) AND ab(assessment* OR test* OR exam*)
Secondary Search Terms
ab("student note*") AND ab(assessment* OR test* OR exam*)
ab("crib sheet") AND ab(assessment* OR test* OR exam*)
ab("memory aid*") AND ab(assessment* OR test* OR exam*)
ab(dictionar*) AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab("graphic organizer*") AND ab(aid) AND ab(assessment* OR test* OR exam*)

[1] For example, one resource aid explicitly mentioned in the RFP was cheat sheets. Although, cheat sheet is a common term, it is colloquial and, in some cultures, has negative connotations, suggesting that the resource should not be used and may not be available to everyone. Other terms that may be used in addition to cheat sheets include fact sheets, information sheets, crib sheets, clue cards, and reference guides.

Secondary Search Terms (continued)
ab("practice exam") AND ab(assessment* OR test* OR exam*)
ab("mock exam") AND ab(assessment* OR test* OR exam*)
ab("specimen exam") AND ab(assessment* OR test* OR exam*)
ab("past exam") AND ab(assessment* OR test* OR exam*)
ab("assignment") AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab("textbook*") AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab("process portfolio*") AND ab(assessment* OR test* OR exam*)
ab("graphic organiser*") AND ab(assessment* OR test* OR exam*)
ab("literary text*") AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab("fact sheet*") AND ab(aid) AND ab(assessment* OR test* OR exam*)

Note. Truncation (*) was used to capture variations of key search terms. References to the search term in the article title or abstract resulted in the inclusion of the article in the initial pool.

The primary search terms produced 554 unique articles and the secondary search terms produced 610 unique articles. Researchers reviewed the abstract for each article to determine the extent to which it related to OBEs or use of specific reference aids. This search established a general sense of the type of information each article could provide. From that initial pool, we excluded articles that were not explicitly related to OBEs or use of reference aids in formative or summative testing.

To ensure that Inflexion researchers were evaluating articles similarly, we completed a two-step calibration exercise in which we reviewed 100 randomly selected articles from the primary search. First, we reviewed 75 articles and met to discuss discrepancies in the articles selected for inclusion. Then, we rated the remaining 25 articles to determine whether the consistency in our inclusion/exclusion decisions improved. Agreement for the first set of articles was 86% and after debriefing, agreement for the remaining 25 articles increased to 96%. The remaining articles were divided among the researchers so that each article was independently reviewed by a single Inflexion researcher. The resulting pool included 149 articles (134 primary and 15 secondary) for inclusion in the literature review. Additional articles were obtained through the backward and forward snowball approach (Wohlin, 2014), in which researchers examined articles that have cited the current article (forward) and examine the references of existing sources (backward) to ensure that all relevant literature has been included. The snowball procedure resulted in additional 32 studies. Although all included articles were reviewed, the literature review narrative focuses on overall themes in the literature, with selected studies used as illustrative examples.

Targeted Organization Search

To ensure the literature review was comprehensive and included the most relevant information available on OBEs, Inflexion researchers conducted a targeted literature review of grey literature posted on the websites of specific organizations. These searches emphasized organizations that

administer assessments similar to IB (e.g., focused on measuring and improving students' higher order thinking), as well as practitioner-oriented outlets that provide practical resources and guidance to teachers. Inflexion used a preliminary list generated during the proposal process and recommendations from the IB research manager and response group to select organizations for the targeted organizational search. In total, Inflexion searched the websites of 47 organizations (see Appendix B for a complete list of organizations) spread across the following seven broad categories:

- Assessment associations
- Companies with assessment-heavy product lines
- Large multinational systems that produce guidance around assessment
- Governments with high-stakes assessment systems
- Universities
- Certification and licensure boards
- Practitioner-oriented outlets

The individual websites of the organizations were searched using the Boolean operator, site:example.com "open book" OR "closed book", which corresponds to the first primary search term used in the academic database search. A total of 210 searches were conducted across the 47 different organizations from across the world. In most cases, only the first 10 hits (if applicable) for each website were scanned and coded for relevance (i.e., yes, no, maybe). After further review of the relevant websites, an initial set of 17 documents (or webpage content) across 9 organizations were added to the literature pool. The 9 selected organizations include 4 assessment companies, 2 certification or licensure boards, 2 universities, and 1 practitioner outlet. Although researchers examined websites from a host of organizations worldwide, the selected organizations were not intended to represent all possible organizations or geographic locations. The purpose of this targeted search was to secure additional literature that may not have been captured by the academic database search and was not meant to be an audit of the extent to which other organizations were incorporating open book assessments into their work. These organizations were selected because their websites contained the most relevant, publicly available information related to open book assessments. Table 2 presents a list of organizations included in the targeted search and Appendix B provides a list of the 47 organizations initially included in the review.

Table 2. Organizations Selected for the Targeted Search

Selected Organizations
American Bar Association
Centre for Teaching and Learning
College Board
INSPERA
NEBOSH
Pearson

Selected Organizations
American Board of Medical Examiners
University of Bristol
University of Oxford

Note. Although the specific purposes of organizations like the American Board of Medical Examiners and American Bar Association do not directly align to IB, both organizations administer high-stakes tests. Thus, some of the general lessons learned about OBEs and reference aids for high-stakes testing could benefit the IB. Further, some of the best research on the impact of OBEs and reference aids on the psychometric properties of the test, as well as exam and item design and testing time limits come from the medical certification field given the legal liabilities associated with that field's certification exams.

The same review process that was used for the academic database search was used for documents and webpage content obtained through the targeted organization search. The main types of documents or webpages included from the organizational search comprised overviews of OBEs and general advice for students, honor code expectations for students completing OBEs, details on security and plagiarism detection measures for OBEs, and guidance for teachers administrating OBEs. The search also resulted in one rigorous study comparing open and closed book medical certification exams. Some of the guidance on OBEs for students and teachers has come in response to the COVID-19 pandemic. The same is true for some of the documents that describe security and plagiarism detection measures. Importantly, the collection of information represented by these four types of documents provides practical guidance and information not typically found in peer-reviewed journal articles. Similar to the role of the cited literature in this narrative, these documents provide illustrative examples of the themes we observed in reviewing documents from all the included organizations.

Coverage of Research Question Topics

Inflexion researchers created a coding scheme to categorize findings by research question and topic (e.g., effects on students' higher order thinking, assessment literacy, reliability). As shown in Table 3, the academic literature search and targeted organization search resulted in adequate coverage of articles related to academic performance, backwash effects, higher-order thinking, student well-being, and academic honesty. Some additional information is needed for psychometric information and teacher professional development. However, the comprehensive literature review resulted in very little information on guidance related to access and inclusion, cultural comparisons, school resources, assessment literacy, and special needs. Information for these topics were culled from the discussion sections of the article and through a targeted Google search. These topics were also explored more explicitly in Phase 2 data collection.

Table 3. Number of Articles Coded to Each Element of the Literature Review Grid

Criteria	Primary Search	Secondary Search	Organization Search	Total
Academic performance	46	11	3	60
Backwash effect	34	8	9	51
Higher-order thinking	25	1	4	30
Student well-being (e.g., anxiety, stress)	12	4	1	17
Academic honesty	4	1	10	15
Psychometric information/Perceived discrimination ability	9	0	2	11
Teacher professional development	7	0	0	7
Guidance related to access and inclusion	1	0	1	2
Cultural comparisons	1	0	0	1
School resources	0	0	1	1
Assessment literacy	0	0	0	0
Special needs	0	0	0	0

Note. The numbers in each column may not total to the number of articles included as some articles are relevant for more than one topic.

Phase 2 Components: Development of Reference Aid Guidelines

Phase 2 addressed the last three research questions and consisted of three primary components: an IB document review, IB staff interviews, and an IB stakeholder survey. These data helped to contextualize the literature review findings, ensuring the resulting information was highly relevant to IB's international education context, and the recommendations and guidelines were appropriate and relevant to IB policies and practices.[2] Each of these components are described in detail below.

IB Document Review

To facilitate contextualization of the literature, Inflexion researchers conducted a targeted review of a limited number of IB documents. The review was ***not*** intended to be an audit of the extent to which IB was currently incorporating OBEs. Instead, the goals of the document review were to (a) explore existing policies and guidelines that may impact the implementation of OBEs and the use of reference

[2] The development of actionable guidelines was heavily dependent on the quality of information available in the literature, as well as specific decisions points for IB, such as the type of reference aids that will be used and the specific context around their use. However, given the fractured body of evidence and the lack of clarity around what OBEs and the use of references aids might look like in the IB, the guidelines were not as specific and actionable as originally envisioned.

aids, (b) describe where OBEs and the use of reference aids could easily fit into existing IB work, and (c) determine if new policies would need to be written or existing policies would need to be adapted to facilitate the implementation of OBEs. Inflexion researchers collaborated with IB project staff to select the most relevant documents for review and secure access to the identified documents. A total of 26 documents were identified for inclusion. Given the exploratory nature of the document review, relevant aspects of the documents were highlighted and qualitatively summarized as they related to the literature review findings. See Appendix C for a list of documents included in the IB document review.

IB Staff Interviews

To solicit IB staff perceptions on specific practices around using reference aids for IB high stakes assessments and classroom assessments, Inflexion researchers conducted 12 group interviews with 32 key IB staff and IB Educator Network (IBEN) representatives. Interviewees covered a range of roles and responsibilities including global research, assessment research and design, MYP and DP subject managers, MYP and DP curriculum development, examiner recruitment and training, educator upskilling, assessment operations, exam preparation, academic integrity, assessment quality, access and inclusion, formative assessment, and IBEN representatives.

These virtual 60-minute interviews were conducted in November 2020. The semistructured protocol had nine items and included probing follow-up questions as needed. Interviewees were asked to provide feedback on their perceptions of (a) OBEs and the use of reference aids, (b) best practices in OBEs and the use of reference aids, and (c) the impact of open book practices on schools and IB work. The IB staff interview protocol is included in the supplemental materials.

Inflexion researchers developed a qualitative coding scheme (Guest, MacQueen, & Namey, 2012) that included overarching *parent* codes that described broad topics and associated *child* codes for specific themes within each broad topic. The fully developed qualitative coding scheme was used to code transcripts from each of the 12 group interviews.

IB Stakeholder Survey

To further enhance contextualization of the literature review data for IB programmes, Inflexion researchers triangulated data obtained through the literature review, IB document review, and IB staff interviews to develop two comprehensive surveys to be administered to IB stakeholders. The *IB staff and IBEN representative survey* included 21 items related to perceptions of open book assessments and the use of reference aids, perceptions of key reference aids and their alignment to specific content areas, perceptions of OBE and reference aid best practices from the literature, assessment design skills, alignment with IB policies and pedagogical practices, and the impact that OBEs might have on their work. The school and IB community perspective was captured in *IB programme coordinator survey,* which included 23 items related to background and experience with IB and OBEs, perceptions of open book assessments and the use of reference aids, perceptions of key reference aids and their alignment to specific content areas, perceptions of OBE and reference aid best practices from the literature, essential teaching and assessment design skills, school resources and assistance, alignment with IB

pedagogical practices, and the impact that OBEs might have on their work. Particular attention was given to ensuring that the research questions were adequately represented on the surveys. Inflexion researchers solicited and incorporated feedback from the IB prior to finalizing the survey. The finalized IB Programme Coordinator Survey was translated into Spanish and French. The IB Staff and IBEN Representative Survey and IB Programme Coordinator Survey in each of the three languages (i.e., English, Spanish, and French) are included in the supplemental materials for this report.

IB Staff, IBEN Representative, and IB Programme Coordinator Sampling

Inflexion researchers worked with IB staff to determine the most appropriate sampling plan for survey administration and to ensure the most suitable representatives were selected for each stakeholder group. For the IB staff and IBEN representative survey, a modified snowball sampling technique was employed where key IB staff were asked to provide recommendations on exam authors and IBEN representatives for inclusion in the survey. The IB research manager selected key IB staff and a sample of senior examiners to participate in the survey. For the IB programme coordinator survey, a random sampling technique was employed by IB staff. An initial list of eligible schools was compiled. Schools offering at least one of the following programmes were eligible for inclusion in the study: MYP, DP, or CP. IB staff randomly selected 500 schools for inclusion in the study. After selection, IB staff implemented a quality control process to ensure the sampling of schools was representative of IB World School strand, region, and program. All programme coordinators at the identified schools were invited to complete the survey. Contact information for the selected IB staff, IBEN representatives, and IB programme coordinators were provided to Inflexion researchers for the purpose of this survey.

Survey Administration Procedure and Response Rates

In January 2021, IB sent a message to the selected IB staff, IBEN representatives, and IB programme coordinators announcing the survey, and in February 2021, the participant pool received an email invitation to complete the survey. Respondents completed the survey online at their convenience at a location of their choice. Respondents were told that the purpose of the survey was to provide information that would help Inflexion researchers develop a comprehensive set of guidelines specific to the use of reference aids on IB summative and classroom-based assessments should IB offer such assessments in future. The surveys took approximately 20 minutes to complete. Table 4 presents the response rate for the surveys.

Table 4. Survey Response Rates

	Total Invited	Total Response	Response Rate
OVERALL	**1225**	**692**	**56.5%**
IB Staff and IBEN	**556**	**372**	**66.9%**
Examiners/Exam Authors	385	263	68.3%
PD Content Developers	28	19	67.9%
IB Staff	143	90	62.9%

	Total Invited	Total Response	Response Rate
IB Programme Coordinators	**669**	**320**	**47.8%**
English LOI* Schools	545	262	48.1%
Spanish LOI* Schools	93	50	53.8%
French LOI* Schools	31	8	25.8%

* LOI = language of instruction

As part of the survey, respondents were asked about number of years they have worked with the IB. Additionally, IB programme coordinators were asked about their experiences in teaching IB courses. See Table 5 for more information on the background and experience of survey respondents.

Table 5. Background and Experience of Survey Respondents

Background and Experience	*N*	% of All Respondents
IB Staff and IBEN Representatives Years of Experience with IB		
Less than 5 years	73	19.6%
5-9 years	59	15.9%
10 years or longer	166	44.6%
I choose not to respond	7	1.9%
Missing Response	67	18.0%
IB Programme Coordinator Years of Experience with IB		
Less than 5 years	50	15.6%
5-9 years	102	31.9%
10 years or longer	166	51.9%
I choose not to respond	2	0.6%
IB Programme Coordinator Experience Teaching IB Courses		
Yes, DP courses (in DP and/or CP)	176	55.0%
Yes, MYP courses	29	9.1%
Yes, both DP and MYP courses	62	19.4%
No	51	15.9%
I choose not to respond	2	0.6%
Legal Status		
Charter	6	1.9%
Private	185	57.8%

Background and Experience	*N*	% of All Respondents
State	121	37.8%
State subsidized	8	2.5%
IB World School Strand		
Private International	68	21.3%
Private National	122	38.1%
Public (ROW)	64	20.0%
Public (U.S.)	66	20.6%
IB Region		
IBA	151	47.2%
IBAEM	100	31.3%
IBAP	69	21.6%

Survey Data Analysis

Survey data were analyzed descriptively to note frequencies and distributions of responses and detect patterns across roles. Open-ended items were coded thematically using the constructs identified during the literature review, document review, and IB staff interviews; in vivo qualitative codes were added, as needed. Respondents who completed at least one question were included in response rate calculations and subsequent analyses.

Guideline Development Working Sessions

In March 2020, Inflexion researchers held three working sessions to develop guidelines around open book examinations and the use of reference aids. Prior to the sessions, Inflexion researchers crafted initial guidelines and recommendations by integrating data gathered from literature review findings, IB document reviews, IB staff interviews, and the IB stakeholder surveys. The three working sessions focused on different aspects of IB work including design and development (curriculum development, assessment design, and general policy considerations); implementation (exam preparation, exam administration, marking and grading, and assessment quality); and professional learning (IB teachers, exam content developers, exam proctors, markers and graders, and IB staff). Each session involved 8 to 12 IB staff.

To the extent possible, Inflexion researchers distilled the information from all sources to provide a comprehensive set of guidelines for the IB. However, given the fractured body of evidence and the lack of clarity around what OBEs and the use of references aids might look like in the IB, the guidelines were not as specific and actionable as originally envisioned. Thus, the working sessions were used to delve into the nuances of each recommendation and facilitate a discussion around how the general recommendation could be further contextualized to IB and how the recommendation may impact the

work of the IB. Finally, Inflexion researchers provided a subjective categorization of each guideline based on the level of importance for successful implementation of OBEs. This coding was designed to facilitate the prioritization of the guidelines and assist IB staff in deciding where to focus their resources. The proposed guidelines are reported in the findings section.

Findings

This study was designed to inform IB's understanding of the current state of the literature around open book assessments and the use of reference aids, both in classroom and high-stakes summative assessments, as well as to explore the best practices in the use of reference aids from a practical perspective, considering the diverse context in which IB operates. The ultimate goal of this study was to develop guidelines specific to OBEs and use of reference aids on IB summative and classroom-based assessments anchored in the best available research evidence and contextualized to the IB. This section presents the summative findings for the open book assessment study, organized by phase.

Phase 1 Findings

A common theme evident in the OBE literature is that the success of OBEs is highly dependent on context. For example, the effects of OBEs on exam performance and students' social-emotional well-being depends on students' preparation, the expectations set by teachers, assessment design, and several other factors. Whether or not OBEs produce more or less stress and anxiety for students also depends on a host of factors. Across all of the areas of research discussed below, past research suggests the interplay of several factors determines how OBEs impact students, teachers, and schools.

As mentioned in the prior section, there are limitations associated with the existing literature on OBEs. First and foremost, a review of literature revealed that the study of OBEs and the use of reference aids has been approached unsystematically, resulting in a fractured body of evidence. Approximately 95% of the research on OBEs takes place at the higher education level. Very little empirical research has examined the effects of OBEs in primary or secondary classrooms, and the majority of the research on OBEs is conducted using small-scale interventions implemented in higher education classrooms. Additionally, very little experimental research exists that directly compares the effects of OBEs and closed book exams (CBEs). The vast majority of research is correlational in nature, which limits what can be said about the cause-and-effect relationship between OBEs and students' exam performance, stress and anxiety, test preparation, and other outcomes. Further, although there is a wide breadth of academic subjects associated with OBE research, only a few subjects have been explored in any depth, most notably, medical studies, psychology, biology, statistics, and mathematics. OBE has only been explored in two or fewer studies for most other subjects. While we can learn from OBE implementation in these contexts, not all findings may translate to the IB context. Moreover, only a few types of reference aids appear to have been studied meticulously, most notably student-created reference aids[3] (specifically crib sheets and "cheat" sheets), teacher-generated handouts, and textbooks. Importantly, much of the research on OBEs does not provide sufficient detail about the nature of the reference aids examined, with many studies simply noting students have access to reference materials.

[3] Where possible, the generic term "student-created reference aids" was used. The terms crib sheet or cheat sheet were used when those particular types of reference aids were being studied and Inflexion researchers were concerned the results may not generalize to other types of student-created reference aids.

With these limitations in mind, the following sections present results from the literature review on the effects of reference aids in high-stakes, summative assessment and classroom assessment. Given the mixed evidence on OBEs, the results presented below focus on trends and patterns between the design, implementation, and effects of OBEs.

Student Performance

Key Takeaways

- Results are mixed regarding OBEs and closed book exams (CBEs) in terms of student performance.
- Teachers, schools, and organizations should not expect to see significant differences in long-term retention between OBEs and CBEs, all else being equal.
- Students will likely perceive OBEs as easier than CBE in the absence of clear expectations. In these cases, students also will likely spend less time preparing for OBEs than for CBEs.
- Providing opportunities to practice OBEs and trainings on OBE study and testing strategies can have a positive effect on future exam performance.
- Some evidence suggests student-generated reference aids may be more effective at improving OBE performance than reference aids generated by teachers or organizations.

Our review of the literature aligns with the only two systematic reviews on OBEs (Durning, Dong, Ratcliffe, Schuwirth, Artino, Boulet, & Eva, 2016; Johanns, Dinkens, & Moore, 2017), which suggest results are mixed with regards to OBEs and CBEs in terms of student performance. For example, in one of the systematic reviews, researchers found that studies addressing examination performance typically found no differences between OBEs and CBEs (Durning et al., 2016). When differences were found, they typically showed students performed better on CBEs relative to OBEs, especially when students spent longer studying for CBEs (Durning et al., 2016).

The central claim of no differences in performance between OBE and CBE can be interpreted as a net effect, meaning some studies find students completing OBEs perform better than students taking CBEs whereas, as indicated above, other studies find the opposite. For example, Gharib, Phillips, and Mathew (2012) studied the differences between CBEs and OBEs (using textbooks or cheat sheets) in initial performance and retention two weeks later in an Introductory Psychology and Introductory Statistics course at an American university. For the cheat sheet exam, students could write as much information as they could fit on both sides of an 8.5 x 11-inch sheet of paper. The results found by Gharib and colleagues are indicative of a pattern observed in studies that find positive effects from OBEs on student performance (e.g., Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008, Nevid, Pyun, & Brianna, 2016, Vogelweid, Kitchel, & Rice, 2014). The authors found students in both open book formats (i.e., access to unlimited resources, permitted student notes) performed better on the initial exam than students who completed a CBE, while there were no differences in retention two weeks later. Taken together, it seems appropriate to suggest teachers, schools, and organizations should not expect to see significant differences in long-term retention on OBEs or CBEs. However, it is also important to consider the purpose of the exam and whether retention of factual knowledge is an

appropriate metric for evaluating OBEs, particularly in an IB context where critical thinking skills are prioritized over factual recall. Regardless, this is often the metric by which OBEs are judged.

There are several factors, some alluded to above, that influence the relationship between OBEs and student performance. One such factor relates to the design of the reference aids themselves as well as the exam format. For example, student-created testing aids (e.g., permitted student notes, cheat sheets, crib notes) resulted in higher test performance than use of testing aids not created by the students (e.g., open textbook exam; Duncan, 2007; Larwin, Gorman, & Larwin, 2013). Researchers speculate this higher performance is attributable to the higher level of engagement that comes from students creating their own testing aids relative to being provided with testing aids (Block, 2012; Larwin et al., 2013). However, a study comparing test performance with and without student-created notes demonstrated that although students performed better on the exam when they used their own notes, the creation of these notes alone is not sufficient for students to learn the material (Dickson & Bauer, 2008).

In an exploratory study to determine actual and perceptional differences in using student created crib sheets for a virology course in veterinary sciences, researchers found that students reported the reference aid helped them to reinforce, remember, and retain course material. Students also indicated they would prefer to use a reference aid on other exams (Vogelweid et al., 2014). However, a two-year longitudinal study on the relationships between the use of crib sheets and student exam performance of veterinary students found the use of the reference aid enhanced student exam performance but did not improve retention of subject matter knowledge. Student survey results suggested students could become overly reliant on their reference aids, thus, decreasing their learning (Rice, Vogelweid, & Kitchel, 2017). It appears that student-created reference aids may be one factor that contributes to better initial performance and long-term retention of material, but not in all cases and certainly not when students use crib sheets in place of preparing for the exam (e.g., learning and developing their understanding of the material).

The design of OBEs may also influence student performance and long-term retention. Heijne-Penninga, Kuks, Hofman, Muijtejens, and Cohen-Schotanus (2012) studied exam practices at three different medical schools, two of which used problem-based learning curricula and the other used traditional curricula. One of the problem-based learning schools used an exam format that included one section that tested core or foundational knowledge and one section that tested backup knowledge,[4] also referred to as two-stage testing. Core knowledge was assessed using a closed book format whereas backup knowledge was assessed using an open book format. The authors expected, but did not find, any differences in backup knowledge between students at the three schools. Further, students who completed the two-stage exam outperformed students who completed traditional open book and closed book exams, both initially and with respect to long-term retention. The authors speculate two-stage testing might allow students "to focus more on acquiring core knowledge, because the body of knowledge they had to know by heart was smaller" and because these students

[4] Core or foundational knowledge is information that students should know and be able to recall without consulting resources. Backup knowledge is the knowledge that a student should be able to understand and apply properly with the use of reference material, if necessary (Heijne-Penninga et al., 2012).

"possibly considered the closed book tests to be more important and were therefore more motivated to study the core knowledge" (p. 493).

Aside from reference aid design and exam format, there are clear indications in the literature that the expectations provided to students influence subsequent performance. Some research suggests students will expect OBEs to be easier than CBEs, presumably because they believe access to reference aids can negate the need for more thorough exam preparation. For example, Jensen and Moore (2009) found students taking OBEs predicted they would have higher course grades than students taking CBEs. The authors assert OBEs created "a false sense of confidence and a less accurate perception of final grades in courses" (p. 61). As will be discussed below, this false sense of confidence often leads to less exam preparation by students (e.g., Agarwal & Roediger, 2011; Christiansen, Lamberg, Nadelson, Dupree, & Kingsford, 2017), which, in turn results in lower levels of long-term retention.

One way to combat a false sense of confidence some students may feel is by providing opportunities for students to practice taking OBEs, either in mock exams or through formative assessment methods. In a study exploring mock exam review sessions (e.g., requiring students to write answers, providing students with the marking or grading scheme for questions) designed to improve student performance on application-based short-essay exams in an undergraduate behavior modification course, students attending mock exam sessions scored higher on exams than students who did not. Student performance increased when the mock exam session included discussion, evaluation, and correction of a sample answer to a question compared to only providing students with copies of study materials about the question (Dotson, 2011).

Opportunities to practice OBEs can also be coupled with specific training for students on how to prepare for and take OBEs. For example, Rakes (2008) conducted an experiment with 122 master's degree students in a teacher education course on instructional technology at an American university. Students completed an online OBE with (treatment group) or without (control) explicit online training on OBE testing strategies. Students in the treatment group had to score a 100% (with retries) on an online quiz specific to OBE test-taking strategies (e.g., Organize Your Information, Manage Your Time, and Prepare the Environment). Students who completed the training scored significantly higher on the course exam than did students who did not complete the training (Rakes, 2008). The authors suggested that students' performance on the OBE may be adversely affected by the differences in the ways they prepare for open book assessments. The findings also suggest "training in open book testing strategies may mitigate the inclination to study less for open book examinations" (Rakes, 2008, p. 5). Additional research is needed on other considerations that may impact student performance, such as the type of reference aids used in the OBE, whether or not the exam is proctored, whether the test is being completed under timed conditions, and any combination of these possibilities.

Student Test Preparation

Key Takeaways

- Students will likely prepare less for OBEs than CBEs in the absence of clear expectations.
- Research is inconclusive on whether students prepare differently for OBEs and CBEs.
- Students' learning orientation (e.g., mastery vs. performance-oriented) will influence how they approach preparing for OBEs. A mastery orientation, which stresses conceptual understanding above memorization and rote recall, is more advantageous for OBEs that measure higher order thinking.
- Students can be effectively taught the study strategies most conducive to preparing for OBEs and how to effectively use reference aids during OBEs.
- Students with weaker study skills may benefit the most from OBEs, especially when these students were provided with explicit study strategy instruction.

The positive effects found in studies where students are afforded opportunities to take practice OBEs or when they are adequately prepared in some other way helps illustrate one common unintended consequence of OBEs: students likely will prepare less for OBEs than CBEs, all else held constant. Durning et al.'s (2016) systematic review found students may prepare more extensively for CBEs relative to OBEs. However, overall mixed evidence was found with regards to the quality of exam preparation, with the authors noting "research exploring exam preparation was equivocal with respect to whether students prepare differently (or at greater length) for CBEs or OBEs" (p. 585).

Student preparation may differ by type of reference aid. There is some evidence to suggest student-created reference aids may lead to better exam preparation strategies than other OBEs (e.g., open textbook, teacher created reference aids) or CBEs. For example, in a study comparing the use of student-created notes (e.g., cheat sheets), open textbook, and CBEs among 198 university students in Ghana, researchers found cheat sheet exams led "to better exam scores, lower pre-exam anxiety and higher knowledge retention than CB [closed book] or OB [open textbook]" (Nsor-Ambala, 2020, p. 48). Interestingly, students studied most for the cheat sheet exam, following next by the CBE, and then the open textbook exam. The author draws on past literature to speculate that students studied more for the cheat sheet exam because the note taking required to create a cheat sheet promotes active learning and the efficient organization and prioritization of information.

Despite the lack of clear consensus in the literature, there are several factors that relate to how students prepare (or do not prepare) for OBEs, or any exam for that matter. As mentioned above, expectations matter. Many students, but not all, will simply not study as much if they expect an OBE to be easier than a CBE (e.g., Agarwal & Roediger, 2011; Christiansen et al., 2017; Jensen & Moore, 2009). For example, in a study of physicians certified by the American Board of Internal Medicine, the authors found that 43% of physicians said the use of electronic resources would reduce how much time they allocated to preparing for the examination (Lipner, Brossman, Samonte, & Durning, 2017). In another study, Myyry and Joutsenvirta (2015) reported that approximately one-third of students allocated less time to preparing for an open book, open web exam relative to a CBE.

One potential reason some students allocate the same or more time to OBE preparation has to do with learning orientations. Many students will naturally approach assessment preparation differently depending on whether the exam is open or closed book. Students who approach learning from mastery orientation will study for conceptual understanding regardless of exam type. Students who approach learning from a performance orientation tend to prepare in a superficial way and are generally uncertain about how to prepare for OBEs (Lindblom-Ylänne & Lonka, 2001). This distinction between mastery- and performance-oriented learning approaches is critical when considering one of the main benefits of OBEs, which is the open book format allows for item types that target higher order thinking skills. Mastery-oriented learning approaches promote building the conceptual understanding necessary for higher order thinking (Gafoor & Kurukkan, 2016). Therefore, it is not surprising that students who approach learning from a deep- or mastery-orientation prefer OBEs rather than CBEs (Karagiannopoulou & Milienos, 2013).

Although the aims and format of OBEs can align with some students' learning orientations, OBEs can also lead to improvements in study skills. A two-year longitudinal study that examined the exam performance of veterinary students who used crib sheets found that students perceived crib sheet use as a way to provide support during studying and testing; however, results also showed that crib sheets could be used as a crutch and could decrease learning (Rice et al., 2017). In a longitudinal study examining the use of open book tests to encourage reading and to assess the improvement of the study skills for students enrolled in first-year college biology courses, students with weaker study skills benefited more from open book tests than students with stronger study skills (Phillips, 2006). The author argues that students with weaker study skills benefited the most from OBEs because these students were provided with explicit study strategy instruction (e.g., reading comprehension tactics, highlighting and marking text, using textbook indexes) and opportunities for practice and feedback prior to and after exams. Students with moderate to strong study strategies likely already had an understanding and experience using these study strategies. This study points to the importance of providing explicit learning strategy training as well as opportunities for students to practice using reference aids, particularly students with weak study and exam preparation skills.

Student Well-Being

Key Takeaways

- Overall, it appears some students report less anxiety and reduced stress from OBEs relative to CBEs.
- OBEs can increase anxiety and stress when students underestimate the difficulty of OBEs, are unfamiliar with the OBE format, or have little experience preparing and using reference aids. Setting realistic expectations and providing guidance is key to reducing student text anxiety and stress.
- Much of OBE research on anxiety and stress is not theoretically grounded and relies almost exclusively on student self-report responses on surveys or in interviews.
- The act of creating a reference aid (e.g., crib sheet) for an OBE may reduce anxiety and stress.
- Eliminating all anxiety and stress from exam preparation may not be beneficial if it results in students preparing less for OBEs due to overreliance on a reference aid.

One of the most commonly cited advantages of OBEs is they reduce test anxiety and stress for students relative to CBEs (Durning et al., 2016). Research appears to confirm this advantage. In study after study, the vast majority of which use self-report surveys, some *but not all* students report less anxiety and reduced stress from OBEs than CBEs. For example, in a survey of 110 university students in Finland, 17% of students reported that open book, open-web online exams reduced their anxiety while preparing for the exam whereas 9% reported an increase in anxiety (Myyry & Joutsenvirta, 2015). Reductions in anxiety were related to students feeling less pressure to memorize facts and having more time to improve their conceptual understanding of the material. Increases in anxiety related to their ability to adequately respond to the different types of exam questions, concerns about exceeding the available time, and worries about not being able to use technology properly during the exam. Similarly, 55% of students felt the ease of response associated with open book, open-web online exams reduced anxiety while 22% felt responding was more difficult, thereby increasing anxiety. Students reporting they felt responding would be more difficult noted concerns related to the use of technology and concerns that the amount of material available would interfere with their ability to write a coherent answer in the given time (Myyry & Joutsenvirta, 2015). Taken together, the results showed that students self-efficacy beliefs may interact with the effect of OBEs on students' stress and anxiety.

Durning and colleagues (2016) note several significant caveats from the available research on OBEs' effect on student well-being. First, the effect of OBEs on student anxiety and stress is often not the primary focus of studies on OBEs. As a result, theoretical perspectives on student well-being are often absent from these studies, making it difficult to fully conceptualize OBEs' effect on anxiety and stress and connect studies to one another across time. Second, nearly all OBE studies that purport to measure anxiety and stress rely on self-report measures where students respond to Likert-scale type or open-ended questions. In these studies, it is generally concluded that OBEs reduce anxiety and stress, "but only a minority of students actually report lower anxiety" (p. 586), thereby masking important differences between students. Finally, and perhaps most importantly, a certain level of stress may not always be bad for students. As Durning argues, "negative emotion like anxiety might actually motivate a student to study for a CBE, which could result in superior performance when compared with an unstressed student preparing for an OBE" (p. 584). Consequently, a careful balance has to be struck between ensuring students have realistic expectations for the difficulty of OBEs and the fact that students can reduce anxiety by avoiding last-minute, stressful cramming that is associated with CBEs. Preparation and practice play a role as well. That is, it seems that anxiety and stress are most likely to be reduced when students are provided with realistic expectations for OBE difficulty and training on preparing and testing strategies for OBEs and have opportunities to practice OBEs.

There is some evidence to suggest that student-created reference aids (namely, cheat or crib sheets) may produce the least amount of anxiety and stress for students relative to CBEs and OBEs (Nsor-Ambala, 2020; Vogelweid, Kitchel, & Rice, 2014; Rice, Vogelweid, & Kitchel, 2017). For example, in an exploratory study to determine actual and perceptional differences in using crib sheets for a virology course in veterinary sciences, researchers found that stress levels reported by students were lower during exams that allowed them to use a crib sheet (Vogelweid, Kitchel, & Rice, 2014). Further, in a two-year longitudinal study to examine the exam performance of veterinary students who used crib

sheets found that students perceived crib sheet use as a way to decrease exam anxiety. Crib sheet use was perceived as positive by students and could be a viable way to combat high levels of anxiety and depression in veterinary students (Rice, Vogelweid, & Kitchel, 2017). One potential reason cheat sheets may produce lower levels of anxiety and stress "could be due to students feeling in control of a critical lever of the assessment process" (Nsor-Ambala, 2020, p. 49).

Backwash Effects on Learning

Key Takeaways

- If OBEs are designed in ways that target higher order thinking skills, teachers may need to adjust their instructional approach to also explicitly target these skills in ways that allow students to see the connection to future assessments.
- If OBEs are designed to target higher order thinking skills, students will have to potentially adjust their study strategies and preparation tactics for exams.
- Providing students with an opportunity to practice OBEs can positively impact student learning and can be an effective supplement to transitional teaching methods.
- Providing students with opportunities to construct OBE questions and engage in discussions with other students around the answers to the questions can improve students' learning and performance.
- OBEs provide students with immediate feedback on gaps in their understanding, a potential, albeit small positive backwash effect.

IB defines the "backwash effect" as "[t]he impact which later parts of a process have on the delivery of earlier parts. In the educational context backwash usually refers to the way teaching and learning is changed by how the candidate is assessed" (IBO, 2019, p. 201). OBEs have two major implications for the backwash effect on learning. First, if OBEs are designed in ways that target higher order thinking skills, teachers may need to adjust their instructional approach to also explicitly target these skills in ways that allow students to see the connection to future assessments. For example, if students are going to be required to use evaluation, synthesis, and analysis skills on OBEs, then teachers need to provide instruction and opportunities to learn that build these skills for students. Given IB's emphasis on higher order thinking skills, teachers should already be emphasizing these skills in their classrooms. Additionally, teachers also need to be prepared to instruct students on how to use reference aids and provide opportunities for practice (Duncan, 2007; Kim, 2019). This may raise concerns related to equitable access to and experience with reference aids. With respect to open-web reference aids, research on IB's onscreen assessments indicated that IB schools worldwide do not have equal access to stable internet service (Jacovidis, Chadwick, Fukuda, Todd, & Thier, 2018). Careful attention must be paid to access issues related to open web as well as other types of reference aids.

Second, and again if OBEs are designed to target higher order thinking, students will have to potentially adjust their study strategies and preparation tactics for exams. Students will not be able to rely on rote memorization or other tactics for basic recall, and instead will need to focus on honing evaluation, synthesis, and analysis skills while also building their conceptual understanding. Again, this should not be problematic for IB given that higher order thinking skills are already emphasized. In

the literature, OBEs do appear to have this type of backwash effect for some students. In a survey of 276 education undergraduate students at a Cyprus university, students reported using higher order thinking strategies to prepare for and complete an OBE but did not employ these higher order thinking strategies on a corresponding CBE. Importantly, the OBE exam included two additional essay questions designed to promote higher order thinking skills, such as synthesizing information from multiple sources (Theophilides & Koutselini, 2010).

Providing students with an opportunity to practice OBEs can positively impact student learning and future test performance. In a study assessing the effectiveness of computer-based self-assessment in medical education, researchers found that performance on the practice exam accurately predicted student performance on future exams. Additionally, use of the self-assessment resulted in better performance in areas in which they initially performed poorly, compared to the areas in which they experience higher performance initially. Students noted the self-assessment was most helpful for identifying gaps in their knowledge and provided them with the opportunities to practice the types of questions that would be on the exam (Leaf, Leo, Smith, Yee, Stern, Rosenthal, Cahill-Gallant, & Pillinger, 2009).

Another tactic for producing positive backwash effects includes providing students with opportunities to construct OBE questions and engage in discussion with other students around the answers to the questions, both of which can improve students' learning and performance. For example, in a study on the use of PeerWise (a student-created online repository of multiple-choice questions), researchers found that writing practice questions has a positive impact on exam performance, but performance is further enhanced by also engaging with discussions with other students around the questions (Duret, Christley, Denny, & Senior, 2018).

Finally, Agarwal et al., (2008) argue that OBEs provide students with immediate feedback on gaps in their understanding, another potential, albeit small positive backwash effect. That is, when students access a reference aid during an exam, it is often due to gap in their understanding or a limitation of their memory. This form of immediate feedback may improve performance; prior research, however, shows delayed feedback after CBEs may be more effective for long-term retention. In Agarwal et al.'s study, OBEs and CBEs that included an immediate feedback mechanism enhanced performance more so than studying material once, repeated study, or CBEs without an immediate feedback mechanism. Both the OBEs and CBEs with immediate feedback resulted in the same level of long-term retention after a one-week delay.

Higher Order Thinking

Key Takeaways

- The little research available is inconclusive with whether OBEs promote higher order thinking.
- Well-designed OBE questions often require students to go beyond factual knowledge to demonstrate higher order thinking, problem solving, application and analysis.
- Two-stage testing that include CBE and OBE sections can be used to target core and backup knowledge, assess lower and higher levels of Bloom's taxonomy, and improve long-term retention.
- Teaching, learning, and assessment format and item design must all be aligned to effectively target and measure higher order thinking on OBEs.

One major tension in the literature on OBEs relates to how researchers define and measure higher order thinking in their studies, if they do at all. Regardless of whether an exam is open or closed book, a key consideration is the learning outcomes that are being assessed. Depending on the course and where it is situated in a series of courses, the assessment will need to target different levels of discipline-specific knowledge, some of which may be well suited for OBEs and some of which may be better suited for other types of assessment. Well-designed OBE questions often require students to go beyond factual knowledge to demonstrate higher order thinking, problem solving, application, and analysis. OBEs tend to focus less on testing students' memorization and recall, which is often encouraged on CBEs. This distinction is one that many researchers fail to make when they compare performance on OBEs and CBEs that include the same items and content.

Despite a lack of specificity in the literature, some researchers have delineated their results by different types of higher order thinking skills. For example, Weber, McBee, & Krebs (1983) compared student performance on a proctored CBE, proctored OBE, and a take-home exam. Each exam contained ten knowledge (lowest level of Bloom's Taxonomy) and ten application (third level of Bloom's Taxonomy) items. For knowledge items, the authors found that students performed best on the take-home exam (a form of OBE), followed by the proctored OBE, and performed the worst on the proctored CBE. However, and more importantly, no significant differences were found across the exam types with respect to the application items.

Similarly, in another study on the relationship between the use of student-created notes and performance on in-class exams, researchers found students performed better on questions at the lower levels of Bloom's taxonomy when their crib sheet contained good information on the topic, but no difference in performance on questions at higher levels of the taxonomy (Hamouda & Shaffer, 2016). There are a number of reasons why this may have happened. It may be that the questions were so well-developed and testing such a high level of knowledge that having references aids was necessary but not sufficient to answer questions. However, at the other extreme, it could mean that the material assessed was not well-represented on the crib sheet. OBEs that allow students to create their own reference aids or select a subset of previously created materials to use during the exam also test students' ability to identify and bring with them the right kind of material—which may or may not be a skill targeted by the assessment.

As mentioned above, one option is to consider a two-stage testing option (Heijne-Penninga et al., 2012). In two-stage testing, proponents recognize there is some information that students should know and be able to recall without consulting resources. This information is considered core or foundational knowledge, also called conversational knowledge in some areas. Given that students are required to have this information memorized, student knowledge is best assessed through a CBE. Proponents of two-stage testing also recognize there is some information that students may not need to know but should be able to apply. This information is considered backup knowledge, or the knowledge that a student should be able to understand and apply properly with the use of reference sources, if necessary. This knowledge is best assessed through OBEs. In this particular study, the authors did not provide information on the types of questions that were included on the open and closed book portions of the exam; however, they were clear that students completed and submitted the closed book portion prior to starting the open book portion of the exam (Heijne-Penninga et al., 2012). OBEs that include a two-stage component may be one way to optimally assess a range of skills and specifically target higher order thinking skills in ways that are obvious and explicit to students.

Exam and Item Design

Key Takeaways

- OBEs can be designed and administered in ways that do not sacrifice psychometric quality.
- Given the same number of items, students generally take longer to complete OBEs, which creates a tradeoff between breadth and depth and ultimately leads to fewer exam items relative to CBEs.
- Without practice, training, clear expectations, and time limits, some students will likely take too much time consulting reference aids in an attempt to identify or construct an answer.
- Well-designed OBEs can enhance the exam's ability to discriminate between low- and high-performers.
- Effective use of resources may be implicitly assessed on OBEs and jeopardize test score validity if this is a skill that the exam is not intended to measure.

The literature review resulted in limited research on exam and item design as it relates to OBEs. Generally, whether the assessment is open or closed book has an impact on the format of the exam and items, and on the psychometric properties of the exam. However, it is generally agreed that validity of test scores is more dependent on the content of the questions than the format of the exam (Durning et al., 2016). However, in the case of OBEs, the format of the exam can potentially affect the number of questions that can be asked. In one of the only systematic reviews of peer-reviewed articles comparing the relative utility of OBEs and CBEs, researchers found that examinees appear to take 10% to 60% longer to complete OBEs (Durning et al., 2016). Thus, if testing time is controlled, it is likely that fewer questions could be included on OBEs, compared to CBEs. All else being equal, this often means we trade breadth for depth and fewer items can be included, which can have implication for validity and impact the test blueprint. Note, the findings related to number of items is most relevant for exams composed of multiple questions and may not apply to exams that include only one item, such as an essay exam. However, allowing reference aids on an essay exam may have implications on the length and/or depth of the essay response that students are able to craft in the allotted time. If students

spend time drawing on the reference aid, then they will have less time to devote to writing their essay. Ultimately, using reference aids takes time and that must be accounted for in determining the number of items and length of time. Realistic time considerations should be based on the number of questions and the type of questions included on the exam.

Further, research has shown that students can spend a significant amount of time consulting their reference aids during OBEs. For example, in a study on physicians using an electronic resource, called UpToDate[5], in both OBE and CBE exam formats, researchers found examinees who completed an OBE took about 40% longer, on average, to complete the exam than those who completed a CBE (Brossman, Samonte, Herrschalf, & Lipner, 2017). Students may spend too much time finding answers instead of applying their knowledge and skills—this can be a threat to validity if students are so bogged down with sleuthing through materials that the exam does not accurately assess their knowledge of the subject matter. Providing testing time limits or limiting the number of authorized reference aids may assist with this issue, if students have spent adequate time preparing for the test beforehand and have sufficient familiarity with the reference aid through training and practice in the classroom.

Despite the potential validity threat posed by longer exam completion times and fewer items, it does appear that OBEs can be designed and administered in a way that does not sacrifice psychometric quality. In one two-stage exam, researchers found that scores on open and closed book sections are strongly positively related. This provides some validity evidence that the open and closed book sections were measuring the same constructs, although more information on the content of the items would be needed to determine whether it was the construct of interest (Heijne-Penninga, Kuks, Schönrock-Adema, Snijders, & Cohen-Schotanus, 2008). Further, in the study of physicians certified by the American Board of Internal Medicine, the use of an electronic resource (UpToDate) enhanced the examination's capacity to discriminate between high and low performers. Researchers noted it was "possibly because able participants used the resource more effectively," (Lipner, et al., 2017, p. 307-8). Thus, effective use of resources may be implicitly assessed on the exam and jeopardize test score validity if this is a skill that the exam is not intended to measure.

Understanding the impact of OBEs on reliability is challenging. In a study comparing the reliability of open and closed book exams, researchers found that the use of open book items in combination with closed book items (e.g., two-stage testing) slightly decreases test reliability of the total exam, but the overall reliability index is acceptable (Heijne-Penninga, Kuks, Schönrock-Adema, Snijders, & Cohen-Schotanus, 2008). Of great importance is recognizing that, in the literature, the variability in scores on OBEs tends to be lower than the variability in scores on CBEs; this decrease in variability directly impacts test reliability (internal consistency) as we are no longer able to differentiate between high and low ability students. That is, if questions in CBE and OBE are both assessing factual knowledge, students who normally would get the question incorrect in CBE have a higher probability of getting

[5] Although the researchers did not explicitly describe the electronic resource, it appears that the reference aid was provided as a restricted, web-based resource, as researchers were explicit that examinees could not access the internet to avoid issues of communicating with others and leaking secure exam content. UpToDate was selected because of its popularity, as more than a million physicians use the resource globally.

the item correct in OBE, making it more difficult to differentiate between those who know the answer and those who do not. This is also a limitation in how OBEs have been evaluated in the literature as most research administers the same exam in OBE and CBE situations instead of leveraging the benefits of OBEs to assess higher order thinking skills. Similarly, it would be difficult to compare the reliability of OBEs and CBEs when different exam items were used. Given that exams with fewer items tend to have lower reliability than exams with more items, reliability of OBEs could be lower than CBEs if fewer questions are asked because of time constraints. Thus, the question of test reliability is less about whether OBEs are *better* than CBEs, and more about whether the reliability is sufficiently acceptable for high stakes testing and optimal for assessing the knowledge and skills of interest.

Assessment Literacy and Design Skills

Key Takeaways

- Very little research exists on the type of assessment literacy and design skills that test authors (e.g. teachers, exam content developers) need to effectively implement OBEs.
- When designing an assessment, test authors should think about the curriculum areas that are best suited to an OBE.
- Test authors and organizations should attempt to use OBE items to target higher order thinking skills.
- Test authors need training on developing and devising effective OBE items.
- Test authors and organizations should be cognizant of OBE length because students will likely be able to answer fewer OBE items that target higher order thinking than on CBEs.

The initial search did not provide much information on assessment literacy and design competency for OBEs. More targeted searches revealed suggestions and guidelines from universities and individual teachers. It is worth noting that, while this information was intended explicitly for teachers, the term "test author" or "exam author" has been used throughout, as anyone involved in exam content development would need these skills. These suggestions and guidelines generally align with test author assessment literacy and design competency best practices that are also relevant for CBEs; however, specific information on the relevance of these suggestions to OBEs are provided. Also note these recommendations are not research-based; however, they can be useful for exploring potential effects of OBEs from experts in the field. Inflexion researchers further probed the necessary skills in Phase 2 of the study.

As is best practice in any assessment, items and exams should be designed with the learning outcomes in mind for OBEs. Test authors should think about the curriculum areas that are best suited to an OBE. For instance, topic areas which ask students to put themselves in the shoes of a professional in their field can easily translate to an open book question, as to replicate real-world conditions students would probably draw on other material, such as industry regulations or professional standards. Assessment items that incorporate reference aids should be used when they are the best choice for assessing the knowledge and skills of interest, not just because it is novel. As

such, it will be important to ensure that test authors have the knowledge and skills to evaluate the appropriateness of OBE alternatives.

Test authors will need to keep the number of questions realistic. In OBEs, test authors might be tempted to give more questions because students have access to resources. However, as previously discussed, the use of those resources takes time. Thus, students may not be able to complete as many items as expected. Further, when time is limited, students will not be able to blindly search through their books or notes for the answer. This will help students to realize that they need to come prepared and have some familiarity with the material or risk running out of time.

As mentioned previously, one major benefit of OBEs is the focus on the ability to assess higher order thinking skills. Writing these types of items is difficult. To do this well, exam authors need to know how to develop and devise effective exam questions that require students to apply their knowledge through analysis and critical thinking. Although writing high quality items that assess higher order thinking is not inherently specific to OBEs, the ability to write these types of items becomes more important as a larger percentage of test items focus on assessing these skills. Additionally, exam authors' abilities to construct questions or tasks that are clearly defined for students becomes increasingly important, as OBEs take longer to complete and fewer questions can be asked. This also helps to limit student confusion and time spent interpreting the question so students can spend their time making use of their textbook or reference aid to effectively answer the questions.

Also, as previously discussed, another major benefit of OBEs is the authenticity of the assessment. Unfortunately, in the literature, reference aids are often allowed without changing the questions on the exam. That does not leverage the benefits of OBE. The key to OBEs is to set questions where students need to do things with the information they have at hand, rather than simply locating it in the text or notes during the exam and reproducing it. Again, although this is also true for CBEs, it becomes more important for OBEs because students have access to information. Examples of these types of items include (a) making use of case-based exam questions that require students to apply critical reasoning skills in response to a scenario; (b) providing students a dataset, artifact, proposal, or idea to interpret, critique or critically evaluate; (c) asking students to provide their own examples of a general principle, concept, or trend; and (d) asking students to come up with some aspect of a design or plan to meet the needs of a specific client, customer, user, or community.

Several organizations provide guides for instructors specific to the guidance above (Centre for Teaching and Learning, 2020; Centre for Teaching and Learning, ND). For example, The Centre for Teaching and Learning at the University of Newcastle Australia (ND) offers a 3-page guide that provides example question stems for OBEs aligned to Bloom's Taxonomy. Although not comprehensive, this guide can help educators see the importance of question stems as well as the type of skills that are required to answer questions that target higher order thinking. Although these guides are useful starting points, there is clearly a need for a more comprehensive approach that includes resources, professional learning opportunities, and other structural supports to build exam authors assessment literacy and design skills for OBEs.

Academic Integrity

Key Takeaways

- Academic integrity and test security are issues regardless of whether exams are open or closed book.
- Accessing unpermitted materials is discussed less than plagiarism and communicating with others (which is most relevant to take home exams and use of unrestricted web resources).
- Clear expectations, honor code agreements, and cheating detection mechanisms (e.g., plagiarism detection software) can be used to combat academic dishonesty on OBEs.
- Student perceptions of whether an OBE would make students more likely to cheat vary considerably.
- To guard against cheating, skillful instructors can design OBE exam items that require the application of higher order thinking skills to generate unique answers; this may make academic dishonesty more challenging and easier to detect.

Academic integrity and test security are issues regardless of whether exams are open or closed book. However, cheating by accessing unpermitted materials is explored less frequently in the OBE literature. The types of academic malpractice that are discussed for OBEs tend to focus more heavily on plagiarism and communication with others (most relevant for take home exams or when unrestricted web-based resources are allowed). For example, one article exploring allegations of academic dishonesty on open book, open note, open internet exams ranged from "inappropriate collaboration to outright plagiarism" (Berrett, 2012). Students may believe they can copy what is in the textbook or notes. Thus, test administrators and teachers must be explicit about when students need to cite their answers and provide examples and instruction on appropriate attribution techniques. As one example, the College Boards' recent *Tips for Success on Open Book/Open Note Exams* (2021) provides general guidance for students on plagiarism issues, including advising students to *apply* rather than *copy* from their notes and being wary of "finding" answers on the internet. In addition, the College Board employed several exam security mechanisms (e.g., plagiarism detection software) during the 2020 testing session for the Advanced Placement program, which was conducted in an open book, open web format due to the COVID-19 pandemic.

When exams employ unrestricted web resources or allow for unproctored administration, it is important to be explicit on whether peer-to-peer communication is permitted. Often students consider their peers to be a resource and may intentionally or unintentionally cheat by discussing exam items with other students (either face-to-face or electronically). As an example, several years ago, Harvard University investigated allegations of 125 students cheating on an unproctored, open book, open web final examination. Nearly half of the more than 250 students in the course "may have worked together in groups of varying size to develop and/or share answers" (Berrett, 2012, para. 3). Importantly, there appeared to be some confusion among students about whether collaboration was allowed or not, despite the instructor claiming the expectations for zero collaboration were clear. Ultimately, students must be provided with very explicit instructions about collaboration during OBEs and what is considered cheating and what is not. Recent research has looked at assessing collective

knowledge and individual contribution to collective knowledge; this topic is beyond the scope of the current study but may be of interest in future research.

One method for providing clear expectations to students is by creating an honor code for exams. The University of Oxford's (2020) *Honour code for open book and closed book remotely invigilated online exams* clearly describes what is permitted and not permitted as well as what types of disciplinary actions that will be taken if cheating takes place. For example, students who access and/or share open book exam questions with other students will face severe disciplinary actions, including possible expulsion from the university. The University of Oxford, like the College Board and e-assessment provider INSPERA, also employ plagiarism detection software. Thus, students must be provided with clear expectations for academic honesty during OBEs as well as an understanding of what consequences stem from unauthorized collaboration, cheating, or plagiarism.

Student perceptions of whether an OBE would make students more or less likely to cheat varied considerably. In a formal survey of 120 Master of Business Administration students from across the world, Williams (2006) found that students were asked whether they thought an open book, open web exam made it more likely students would cheat. About half of students disagreed that open book, open web made cheating more likely, about one quarter remained neutral, and one quarter felt it cheating was more likely. The author argued, however, these data simply correspond to regular patterns of cheating, meaning students who cheat will cheat regardless of the exam format (Williams, 2006).

Further, one article postulates that skillful instructors can make cheating and abusing the online testing environment (e.g., copying questions, sharing, and taking screenshots) a challenging, time consuming endeavor for students. However, the articles do not provide much information on *how* to do this. The general consensus is that cheating can be reduced by including questions that require analysis and synthesis rather than recall questions answered through memorization or consultation of a reference aid (Lam, Williams, & Chua, 2007; Williams & Wong, 2009). For example, in Williams and Wong's study,

> "Students felt that the OBOW [open book, open web] exam [was] designed in such a way that cheating is difficult given the exam is based on a current case study that was developed and customised, taking into consideration the theories, concepts and issues covered in the discussion boards and assignments" (p. 231).

School Resources

Key Takeaways

- Very little research exists on what type of school resources are needed to support and train teachers to effectively implement OBEs.
- Key considerations revolve around providing time and resources for students and teachers to develop reference aids, providing equal and equitable access to reference aids (e.g., printed materials, digital devices), and developing systems to potential restrict access during OBEs (e.g., restricted web access).

The initial search did not provide much information on school resources necessary for OBEs. However, school resources could be gleaned from discussion and recommendations in the reviewed research articles. It is important to note these recommendations are not research-based; however, they can be useful for exploring potential effects of OBEs from experts in the field. A preliminary list of resources that should be considered for OBE is provided; Inflexion researchers further probed the necessary school resources to successful implementation of OBEs in Phase 2 of the study.

- *Student-created reference aids:* Students will need time to create the materials and explicit guidelines on what is allowed. Similarly, if the reference aids are teacher-created, the teacher will be required to invest time to create the reference aids and the quality of the reference aids may vary considerably among teachers.
- *Print reference aids:* Students will need a clean copy of the textbook and/or printed materials. This may increase printing costs for schools or for the IB. Depending on the reference aid, it may also affect the number of copies of specific textbooks, literary texts, or dictionaries that the school must have available. Further, students will need desk space to accommodate the print materials.
- *Electronic references aids:* Students will need their own personal device or access to a school device and access to outlets. This may not be an issue if the exam is digital. Additionally, given that many students have multiple technologies, it is difficult to restrict access. Students' computers may be restricted, but they may be able to access their tablet, phone, or smart watch to access restricted or blocked materials.
- *Restricted access to reference aids:* The teacher or proctor will need to check students' references aids to ensure they are only using authorized materials or work with the technology team to limit digital access. Given we live in an era of ubiquitous technology, it is difficult to completely restrict access. And allowing students to use some resources but not others becomes more difficult to manage. For example, if students are allowed to use open access web resources, it is important to ensure that students are not communicating with each other. This becomes difficult to enforce—it is possible but requires additional considerations.

Phase 2 Findings

The following sections presents a summary of the data used to contextualize the literature review findings. First, a synthesis of the IB document review is provided, followed by a summary of the IB staff interview results and the IB stakeholder survey results. This section concludes with a summary of the working session discussions around the development and contextualization of guidelines for OBEs and the use of references aids in the IB context.

IB Document Review Synthesis

The IB document review involved drawing a connection between IB documentation and the OBE literature and then expanding on how OBEs may translate to the IB context.

Connecting the Dots Between IB Documentation and the Literature

Key Takeaways

- The goals of implementing OBEs in the IB context include allowing for the assessment of a more expansive set of higher order thinking skills, producing positive backwash effects on teaching and learning, and reducing student anxiety and stress.
- To realize these goals, a number of factors must be addressed and accounted for, including designing assessments to target higher order thinking skills, providing clear expectations around the use of reference aids, and providing teachers and students training and practice with OBEs.

The best-case scenario in the IB context is that OBEs: (a) measure a more expansive set of **higher order thinking skills** associated with the IB Learner Profile, (b) **produce positive backwash effects on teaching and learning**, and (c) **reduce the anxiety and stress** associated with summative exams. That is, OBEs have the potential to measure skills more closely aligned with IB's mission in a way that improves teaching and learning and students' social emotional well-being. However, perhaps the most critical takeaway from the existing literature on OBEs is that a number of associated factors must be addressed and accounted for if this best-case scenario is to be realized.

The list of factors that affect whether the potential of OBEs is achieved is extensive, and we know more about some of these factors than others from existing research. That said, a reasonably clear picture of how these factors are related is visible when collectively examining the existing literature on OBEs. First and foremost, OBEs must be **designed** in ways that target specific higher order thinking skills. Without this initial step, the only major advantage of OBEs over CBEs would be a possible reduction in student test anxiety and stress. If the only change in exam design is the fact it is open book, research suggests an unintended consequence will be students spending less time preparing for the exam itself. A second and related design issue is that the specific reference aids used in OBEs should align to the specific higher order thinking skills targeted by exam developers. For example, exam questions that require students to seek out sources from the internet to answer an open-ended theory of knowledge question are testing students' abilities to analyze, evaluate, and synthesize information as well as their web searching skills, the latter of which could introduce construct irrelevant variance (i.e., knowledge or skill that affects scores but is not part of what is being targeted on the assessment). Third, research shows that testing students' higher order thinking skills in complex, challenging questions that require use of reference aids will lead to longer testing times for students, which may require fewer overall items. Psychometric testing will be needed to determine the relative impacts on validity and reliability in these cases. Finally, another key design consideration is access to and experience with various reference aids. Students may have different levels of access to certain reference aids (e.g., unrestricted web) based on geographic, socioeconomic, and individual (e.g., learning disabilities, language difficulties) factors. Failing to account for these factors will affect the validity, reliability, and fairness of OBEs. This short list of design issues is not meant to be exhaustive, but it does illustrate that for the three main advantages of OBEs to be realized the design of the exam needs to take these issues into consideration.

With a well-designed OBE in hand, the next step is to set clear **expectations** for students and teachers. Evidence shows that if students are not provided with an understanding of the design and aims of OBEs, they will likely spend less time preparing for the exam and feel overconfident in their ability to perform well. In these cases, although the OBE may target higher order thinking skills, neither the positive backwash effects on learning nor improvements in students' social emotional well-being will be realized (e.g., unprepared and overconfident students may experience greater anxiety and stress during an OBE). Additionally, unclear expectations will also likely result in intentional or unintentional instances of academic dishonesty in the form of plagiarism or unauthorized collaboration between students. Thus, clear expectations can serve several functions, including providing students with the type of preparation and testing strategies that are useful for OBEs, signaling to teachers what instructional and classroom assessment practices align with the intent of OBEs, and specifying what type of behavior is acceptable from an academic honesty perspective. Several organizations have produced documents that include these and other types of expectations for students and teachers.

A well-designed exam coupled with clear expectations for students and teachers is necessary but not sufficient for fully realizing the major advantages of OBEs. Both teachers and students need **training** and **practice** with OBEs. First, teachers need training and opportunities to practice developing classroom-based formative and summative OBEs that target higher order thinking skills. Additionally, teachers need training and an understanding on how instructional practices and routines can reinforce a focus on higher order thinking skills, thus providing greater alignment between teaching, learning, and assessment. For example, teachers should be trained on how to create quality, equitable reference aids themselves. Teachers should also instruct students on creating and using self-generated and other reference aids. Evidence shows this type of practice and training for students, provided by their teachers, can help familiarize students with the OBE format and reference aids themselves. When this training is equity-oriented, it can help reduce variation in test performance that is due to differences in experience with reference aids across students. Finally, more experience and familiarity with OBEs will reduce overall test anxiety and stress for students.

Keeping the central goals of OBEs—measuring higher order thinking skills, producing positive backwash effects, and reducing test anxiety and stress—at the forefront of any implementation effort along with addressing and accounting for the exam design, student and teacher expectations, and training and practice needs is only a starting point. As the following section describes, how implementation unfolds is highly dependent on the unique IB context.

OBEs in the IB Context

Key Takeaways

- From a policy, standards, and practices standpoint, it appears very little needs to be changed for OBEs to fit within IB's programmatic system.
- One of the first steps in planning for the implementation of OBEs is to determine what reference aids are most appropriate for specific IB exams and identifying what tradeoffs may exist.
- Implications for access and inclusion on OBEs will be key and must include considerations around access to different reference aids, as well as geographic and socioeconomic considerations, given that IB is active in 158 countries.
- As with any change to assessment, IB will need to examine the impact of the transition to OBEs on assessment quality, specifically in terms of reliability and validity.
- IB will need to provide documentation that clearly describes expectations to all stakeholders.
- Teachers will need professional development to support the effective implementation of reference aids in IB exams.

The good news is that IB's current mission and goals for students, teaching and learning philosophy, and assessment principles and practices all align closely with the central motives for OBEs. From a policy, standards, and practices standpoint, it appears very little needs to be changed for OBEs to fit within IB's programmatic system. For example, IB summative assessments are already designed in part to target higher order thinking skills and produce positive backwash effects on teaching and learning (IBO, 2019). OBEs can represent an incremental step toward achieving these objectives, rather than requiring an extensive overhaul of existing systems. Developing a clear rationale and framing for the purpose of OBEs in the IB context will be critical for securing buy-in from the stakeholders who will be tasked with implementing new systems and practices to support OBEs. A shared vision for OBEs also will help stakeholders understand the connection between the different factors that make or break the success of OBEs.

When considering the IB context, careful attention must be paid to differences among subject areas and courses and to issues of access and inclusion. For example, the IB Diploma Programme includes six subject areas and more than 50 individual courses with associated exams. OBEs should look different in each of these exams based on the specific higher order thinking skills being targeted, the nature of the content being tested, and the type of reference aid(s) that would be most appropriate. Existing IB assessment policies provide a starting point for developing exam-specific needs for OBEs. For example, *The conduct of IB Diploma Programme examinations* (2018) shows that some Diploma Programme exams already allow certain reference aids whereas others do not. The most common reference aids appear to be formulae sheets, data booklets, and dictionaries. More generic reference aids, such as crib sheets and open-web access, are likely to be more appropriate and relevant to certain exams and not others. Thus, one of the first steps in planning for the implementation of OBEs on a wide scale is to determine what reference aids are most appropriate for specific IB exams and identifying what tradeoffs may exist among different reference aids.

An equally important step is identifying and addressing issues of access and inclusion as they relate to OBEs. Given IB is active in 158 countries, issues of access and inclusion relate to geographic and socioeconomic considerations. For example, the vast majority of research on OBEs has been situated in the United States and Europe. Another geographic consideration relates to access to reliable internet for open book, open-web exams. Students from economically challenged backgrounds may not have as much experience with using the internet for research purposes as their more affluent peers.

On an individual basis, IB provides specific assessment principles and practices for students with (a) learning disabilities; (b) language difficulties; (c) specific learning difficulties; (d) communication and speech difficulties; (e) autism spectrum disorders; and (f) social, emotional, and behavior challenges. The *Candidates with Assessment Access Requirements* (2017) document provides specific guidance on inclusive assessment arrangements for these groups of students, such as modified exams, additional time, access to a reader or reading software, access to interpreters or augmentative communication devices, and access to different types of calculators. Different types of OBE formats and reference aids will undoubtedly affect inclusive assessment arrangements in different ways. In particular, careful attention should be given to ensuring whatever advantages are produced by access to different reference aids are experienced equally by all students. For example, if IB employs crib sheets within an open book format, Braille transcription services should be provided to students who are blind. There are numerous other examples. In the end, failing to account for inclusive assessment arrangements will threaten the validity, reliability, and fairness of OBEs.

Several additional IB systems and policies will need be modified, and in some cases created, to support implementation of OBEs across IB exams. First, the inclusion of reference aids, which will necessarily look different across IB exams, could potentially require modifications to exam format and items. The IB Assessment Division, including the Assessment Research and Design department will need to carry out psychometric testing to ensure any changes stemming from shifting to OBEs results in acceptable levels of validity and reliability and new exams are fair from an access and inclusion standpoint. The IB Assessment Principles and Practices department also will have to ensure any changes are accurately represented in policy documentation so that schools and teachers are adequately prepared and informed about the effects of shifting to OBEs.

IB will need to provide documentation that clearly describes expectations for teachers and students related to OBEs. These expectations should provide a rationale for OBEs, the advantages of this type of exam format, and key considerations for teachers and students. As mentioned above, these types of expectations are critical for guarding against some of the potential unintended consequences that can stem from OBEs (e.g., lack of preparation, overconfidence). Finally, teachers will need professional development (e.g., training videos, documentation, workshops) to support the effective implementation of reference aids in IB exams. The types of teaching and assessment design skills this type of professional development could target includes but are not limited to designing classroom-based OBEs, developing OBE items that target higher order thinking skills, generating teacher-generated reference aids, training students on creating quality reference aids, and leveraging OBEs to produce positive backwash effects on teaching and learning.

IB Staff Interview Summary

Inflexion researchers conducted group interviews with IB staff to better understand their perceptions of specific practices around using reference aids for IB high stakes assessments and classroom assessments. Information from the interviews informed the development of the IB stakeholder survey and allowed for the contextualization of the literature review. Interview results were also key in developing initial guidelines for OBEs and the use of reference aids in IB's context. The following section presents a summary of the IB staff interview findings.

Experience with Open Book Assessment and the Use of Reference Aids

Interviewees were asked about their experiences with open book assessments and the use of reference aids. Most interviewees reported having some experience with OBEs and reference aids. Their experiences generally revolved around taking OBEs as a student, their children taking OBEs as students, administering OBEs as an educator, and experiences with reference aids that are currently being used within IB.

IB staff generally shared positive perceptions regarding the use of reference aids in formative and summative assessments. As one staff member noted, "My general perception is that they should be welcome especially since we're kind of doing a shift in the educational landscape towards the ability to do open problem solving." Other staff commented more broadly saying, "I'm a fan of open book, definitely," "I'm super open, ready for it, love it, support it all the way," "I would just say that I'm very open to it. I think most of the colleagues and DP teachers that I've worked with over the years in the humanities are very open to it," and "Make it work. We know it's got pitfalls, but make it work." None of the interviewees expressed negative perceptions of OBEs and the use of reference aids.

Specific Reference Aids for Specific Content Areas

Inflexion researchers also explored IB staff perceptions of specific reference aids and their use within specific content areas. Although interviewees agreed that specific reference aids were better suited for specific content areas, it was difficult for them to say which reference aids were best suited for which content areas. As one interviewee stated, "It needs to be looked on subject-by-subject basis. As for specific examples, without being given a specific subject it's very difficult to come up with a specific example." Another interviewee added, "You cannot say we're going to allow one particular type of reference aid no matter what the subject is… The type of aid that is allowed is very closely linked to the subject or topic area and the assessment objective."

A few respondents focused on the appropriateness of certain types of reference aids. As one interviewee shared, "There's just so many subjects… It's always easier if it's more restricted." The interviewee went on to say that that IB-created reference aids would be the best approach to allow control over the reference aids and standardization across schools and students. However, this becomes more challenging when the reference aids are not IB-created. A few interviewees expressed concern about the logistics associated with restricted access saying, "I don't think we can say, 'You can bring in this, but you can't bring in that.' Because we would never be able to check it all in the time"

and “Normally there would be regulations and guidelines of what you would be allowed to bring in… I don't understand how we could possibly police that, working the way we currently do.”

Other respondents commented more broadly about how reference aids would be better suited for certain IB subjects, although they did not specify the type of reference aid.

> *I'm thinking ESS [Environmental Systems and Societies] would be a really good subject to do this for because the world is constantly changing, environmental viewpoints are changing. IT and computer ethics would be another one where you have got constant change like, the role of social media.*
>
> *One of the subjects that this would actually be most appropriate for is ITGS or Digital Society. And it's currently in review… And that may be a starting point to actually pick something that's in review, it can be looked at. It's small enough that it can be trialed. You could work around what comes out of the trial.*

Similarly, some interviewees preferred to think about the types of questions that would be asked.

> *I would say it depends on the type of assessment you are giving, because if you are giving an assessment that is essentially a recall type of assessment, then obviously it's pointless… We've got to move always to higher level thinking questions that are at minimum on the sort of apply type of level. I think IB examinations are designed very often to give access to all levels. So there are obviously very straight forward ones. [In OBEs], there is really no point in that.*
>
> *I think it comes back to the different nature of the subjects. For example, Computer Science. You do want the kids to be able to recall information they should understand… You shouldn't be doing a computer science job and having to Google what the basics of the job are… In the more applied subjects, the content knowledge is less critical. It's about how you apply that knowledge and how you argue and justify… There the open book approach is much better.*

A couple interviewees were able to put the two together to provide more specific suggestions.

> *I think dictionaries would probably work very well in language and literature but probably not language acquisition because we are assessing different skills.*
>
> *I can sort of see how in things like literature, you probably want, you might want books for them so they can pull out quotes that are particularly useful.*
>
> *I think in subjects like business management, economics, things like a glossary would be useful... And even in subjects like history, thinking back to my own DP history exam, recalling set dates of events, and being concerned about tripping up on those, I think would also be very helpful.*

Some interviewees were apprehensive about the idea of allowing different reference aids for different subjects. One IB staff member characterized it by saying, "I can imagine it being really hard if it's going to be different for different subjects. If you're saying, 'Well, in history it's fine to have anything you like, but geography you can only bring in half a page of facts,' that's just asking for trouble, isn't it?" However, it's important to note that IB already allows different reference aids for different subjects (e.g., an information booklet in science, a calculator and formula booklet in mathematics, and dictionaries in some of the languages). Thus, this may not be as challenging as this interviewee anticipates.

The Purpose of Assessment

A key theme throughout the IB staff interviews was the importance of understanding what was being assessed with OBEs. Interviewees were quick to state that the use of OBEs and reference aids depended on the skills that the assessment was targeting. Interviewees agreed that OBEs and the use of reference aids were well suited for assessing higher order thinking skills but were less useful or appropriate when the assessment was targeting declarative knowledge.

> *Given that there are learning objectives and assessment objectives in each program, it's very clear that we value some sort of content knowledge, in order to achieve those objectives… It's very clear that we're testing some sort of levels of knowledge from the concrete to the abstract… And that the exams as we write them are supposed to be reflective of that.*
>
> *I think that really depends on what is being measured and what the desired measurement is supposed to entail for that particular examination.*
>
> *If it's a purely, recall at a knowledge level and it's open book, then what's the point?*
>
> *I'm very open to the idea of it, but I've yet to actually find a case where we need them to assess what we want to assess. If someone says they want to do an open exam, we say, "Okay, what are we trying to assess here, and how do we want to do it?" Then if it becomes obvious that we absolutely need one, then, or not that we absolutely need one, that it's beneficial, then absolutely, but a lot of the time I don't see it.*

Another IB staff member added, "Summative high stakes have different pressures in terms of outside certification. So, it has to be treated cautiously given the context and nature of what is being assessed, and for why, and for whom."

A few interviewees also expressed concern that OBEs may be assessing skills that are not intentionally being targeted on the exam, such as the ability to create or use reference aids, time management, and use of technology, for certain reference aids. As one staff member noted, "I think we have to be careful that we don't end up assessing one skill that we don't intend to assess when we're assessing a different skill." Another interviewee added, "If you have an open book, you're really just measuring to what extent a student is able to create a reference aid and less about whether they were able to use that reference aid to display any type of understanding." A third interviewee stated,

> *Anytime you use the internet in OBEs, you have hidden competencies that you're testing… time management, you're not just testing their media literacy, but their literal ability to search a screen or to search a technology to find something.*

Alignment with IB Philosophies

Most interviewees noted that OBEs and the use of reference aids aligned with IB philosophies towards teaching, learning, and assessment. Many of the comments about alignment with IB philosophies centered on the authenticity of the exam and the ability of OBEs to assess higher order thinking skills.

> *Philosophically, I would like to think that it aligns closely with IB like assessment philosophy. We always want to assess the high order thinking, and that also aligns with the way we want to learn and teach in classrooms.*

> *I think they align perfectly… we want to test concepts and actual learning rather than just recall... That's one of the key areas, is that we want a holistic education rather than just recall and that's one of the reasons why people chose the IB.*

> *I think an open book exam, it's quite attractive from the organization's point of view because it does represent authentic experiences, [the] sort of work that they're going to be doing in the real world in the future.*

Two interviewees commented on the extent to which OBEs align specifically with IB's Learner Profile.

> *I think about the learner profile… I don't think it says anywhere, "is able to recall facts without reference to a book." It's not really focused on your ability to learn things by book and then spout out in an exam. It's all about risk taking and being compassionate and that kind of thing.*

> *I would like to see our assessment moving more towards the learner profile in a more forceful way than it is at the moment. I think we expect these things to happen in the classroom and that's how the IB learner profile is actually developed, but I would like us to have some evidence of the IB learner profile working in action… The learner profile talks about being reflective, open-minded, taking risks, taking care, principle, balance, inquiring, communicative and think these are not things you can demonstrate if at the same time you have to memorize facts.*

Further, one IB staff member noted the use of OBEs and reference aids in the DP could affect the vertical alignment among IB programs, stating, "DP exams could change in nature and would become perhaps more like the MYP. So by allowing [OBEs], you would potentially strengthen the alignments across MYP and DP, and in the DP subjects of CP."

A number of IB staff members provided examples of how IB assessments already include some components of OBEs and the use of reference aids. IB staff noted there are permitted reference aids, including data and formula booklets in mathematics and sciences, and the use of a dictionary in classical languages. IB staff also mentioned the pre-seen case studies that are released prior to the

exam and taught in class, as well as assessments, such as the extended essay and reflective project, where students complete the assessment with access to any materials they wish.

Impact of Open Book Practices on IB Work

Overall, IB staff did not seem to think that the implementation of OBEs and the use of reference aids would impact their work significantly. One interviewee characterized it by saying, "I don't think it will have a detrimental impact. I don't think even the most negative scenario would not be so challenging that it would not have been worth it." A few interviewees noted it might be more work; however, it was still worthwhile. As one IB staff member stated, "It could affect our job by making it more complex, but also maybe more interesting." Two other interviewees added,

> *No matter what we do, it will involve a lot of work. What that work looks like depends very much on the decision made for each individual component on each subject. It will not look the same, basically, but it will definitely be more work and a lot of it. But most likely, more than worthwhile.*

> *Assuming they're not mandated, it would be like another option in the toolbox of assessments you can use, and ways you can vary an exam to suite its purposes… It will basically free up quite a lot, I think, of rules in terms of designing assessments.*

Although IB staff generally did not see a direct impact on their particular duties and responsibilities, many interviewees shared where they thought OBEs would have the biggest impact. Overall, most interviewees held the opinion that the implementation of OBEs would have the greatest impact on marking and exam authoring. Interviewees noted the biggest challenge for authors would be training them to write question that align with OBEs. Additionally, three IB staff members thought that some of the more conservative examiners and authors may resign if they are asked to participate in OBEs. Interviewees also noted, because OBEs take longer to mark, more examiners may be needed. In addition to the increase in time for marking, the subjectivity of marking is also likely to increase. One staff member stated, "The only problem I see, is that if we were trying to assess more higher order skills, then that's going to be far more subjective, they're just going to make it more difficult for the examiners to potentially mark to the standard of the principal examiner, so I think we might have a higher failure rate for examiners, certainly initially." Finally, four interviewees commented on the impact OBEs would have on the criteria to which markers are marking and how that might impact examiner training. The following comments illustrate their sentiments.

> *[Examiner] training would need to be different. I think follow through marking would be rather important if we're expecting the candidates to draw a conclusion that they've been given. I don't particularly think it's fair that just because they came up with the wrong answer, we don't continue marking. I think follow through marking would need to be involved.*

> *So as I said before, it would be a complete change in the way examiners work and what they would assess. I assume that from marking perspectives, we would stop asking them to*

> *focus that much on knowledge and focus more on the other strands that we assess. It would involve training for all examiners that we have on our books for all subjects affected.*

Another interviewee highlighted, "It's not any different than from when an assessment gets a new assessment model. It wouldn't be different. It's a new way of marking every five [to seven] years… and we go through it."

IB staff also noted some logistical issues related to the preparation time for assessment, saying "In terms of assessments, our assessments are in advance 18 months before they're out there. The more sophisticated they're going to be, the longer the timeline is going to have to be as the lead in." Additionally, interviewees also commented on the implications for assessment operations, particularly around producing the materials in multiple languages:

> *It's another element in the process that needs to be absolutely right and quality controlled to the same extent that the assessments are. That takes people and time and energy and money and expertise, not that we don't have or have access to, but we're going to see more of it.*

IB staff also emphasized the need for professional learning and support for teachers. Although most of the comments were broad, it was clear that IB staff thought teachers needed professional development on and high-quality examples of OBEs and the use of reference aids. One IB staff member suggested that OBEs should be part of the subject workshops for courses that include OBE, rather than just being a separate workshop. Interviewees also noted that teachers will need additional support around designing good classroom assessment for OBEs, and specifically training on how to write high-quality OBE questions. One interviewee suggested that IB rethink how programme coordinators are used or encourage schools to add coaching positions that work directly with teachers. Additionally, IB staff agreed that teachers would need guidance on how OBEs would affect teaching and learning. One interviewee summarized it by saying,

> *To be honest, I do not see it as any different from our normal curriculum, when we changed the assessment models and teachers have to adapt to that and teach the students a different way to make sure that they are adhering to and giving the kids the best chance to succeed with the new assessment model. It would be a change of practice but there's a change in practice in every single subject every five (sic) years anyway.*

Benefits of Open Book Exams and the Use of Reference Aids

Throughout the interviews, IB staff noted many benefits of OBEs and the use of reference aids. The most common benefit of OBEs was the authenticity of the exam. Generally, interviewees agreed OBEs and the use of reference aids more closely approximate what would be expected of students in their future careers. The following comments illustrate the sentiments shared by IB staff.

> *I feel like it needs to mimic real life. We don't go around our lives without researching things, that's not what is expected in real life is to Google the necessary books that we*

> *might need to get our job done properly. I think if I did the job badly, I said, "I didn't use any book." I think my boss wouldn't be happy with me.*
>
> *I believe that life is open book, so I think that assessments that are not lack a considerable amount of authenticity. That said, I acknowledge the logistical issues that that introduces.*
>
> *I think it is interesting because we are currently in an era where you don't need to memorize things. Now it's just right at your fingertips. It's how you access it now is more key than what you do with that knowledge.*

Another common benefit of OBEs and the use of reference aids, already captured previously, was their ability to assess higher order thinking skills.

> *I think it will be better for the assessment because then they can focus on actually testing for the students' knowledge as a concept, rather than how well they remember fact or a book.*

Academic Honesty in Open Book Exams

Although staff were overwhelmingly positive about OBEs and the use of reference aids, many of the staff members expressed concerns about the implementation of OBEs and the use of reference aids in IB's context. One of the most common concerns related to academic honesty.

> *One area where things would need to be considered, would be in the assessment procedures and the policies surrounding academic honesty. That's where either very fine lines have to be drawn about what is and isn't allowed and how those tools and resources are allowed to be accessed.*
>
> *How do we, especially students writing notes and all that, then how are we able to ensure that they are producing their own work and they're not getting help from other people. I think that those things are grey areas that needs to be resolved first before we ensure a fair assessment environment for everyone.*
>
> *One of the things is you always have that one student who is the over achiever and might be selling their notes and nobody else is taking their notes. That can be an issue. Is that really testing the student's knowledge or their ability to just take that piece of paper and put it out?*
>
> *Anything that goes into the examination hall, or room represents a risk and the sense of students using it to cheat.*

Interviewees also expressed concern about the commercialization of the reference aids. As one interviewee shared, "You can guarantee that help sheets will appear on the internet, in terms of how to maximize what you can get out of these notes, probably for sale, which I suppose that happens." Two other interviewees commented:

> *It's the way that for-profit business always gets in on these things… One thing I worry a little bit might be the creation of resources to take into an exam, and that might be why I would be more inclined to keep it restricted, because otherwise you are going, I am fairly sure, you are going to have business stepping in and offering things that may not be helpful or equitable.*
>
> *Whenever we're packaging anything about IB we run the risk of commercialization in a competing vendor world. And we see at our conferences rows and rows and rows of displays where third parties are competing to sell everything from textbooks to exam aids to manipulatives.*

Equity in Open Book Exams

Most interviewees expressed concern regarding the equity of OBEs and the use of reference aids, specifically around information available to schools across countries and in different languages.

> *I'm going to jump in now with a warning message that we really should be aware of. I'm really keen on this, but within the context of the IB, we've got the three working languages plus the other languages now that our subjects are offered and it's not a level playing field in terms of even books, let alone websites and so on that they have access to.*
>
> *Language equity would be hugely affected by OBEs, if they're completely open. Because we would have to be very careful about designing questions that would allow for access in all of the languages that students can take the examinations in. And definitely quality of resources and the ability to find resources is affected by its language and location.*
>
> *In the global nature of the IB, we would have to make sure that all of the resources were available in all the countries, but also obviously that the translated versions were also the same standard but I think you'd automatically disadvantage students… We couldn't expect ourselves to produce resources for every single language and equally we couldn't guarantee that schools would produce resources to the same level.*

Some interviewees discussed equity in terms of disparity among school resources.

> *When we talk about OBEs, I do question equity, because what happens if a wealthier school has a real quality book and another school is using a guide that's maybe not as robust?*
>
> *One of the things I'm hugely aware of is the different resources that [students] have available to them. If it's up to schools and students to decide what the kids can take into an assessment, I would be very worried because I think that would leave a massive difference of opportunity.*
>
> *If you're going to have searching in an exam, you're then going to be dealing with the infrastructure of the school. And schools which do not have good infrastructure, the kids are going to be instantly disadvantaged because they're not going to be able to get as*

> *much information as quickly as they need. Or the firewall of the school, or the policy whether the schools in the district and the district has a blacklist or a whitelist.*
>
> *If there is state-funded school, it doesn't have access to the big funds that the private schools have access to. If they've got a book that was say the first edition of a textbook that can be allowed in. But the other students are using the third edition of the textbook, then do they get an unfair advantage from, over the other students even though this probably is not their fault.*

School resources are an important consideration for OBEs and the use of reference aids. The availability of resources will always be a struggle for many schools. This can become challenging for schools. If the reference aids are digital, it may be a challenge to ensure that all students have access to a computer, although this is less problematic with the transition to e-assessments. If the reference aids are hard copy, it may be resource intensive for schools to print the reference aids for their students.

Some IB staff focused on equity related to students' abilities to create reference aids. As one interviewee noted, "Some students are quite good and already know how to take notes. Some students reach the end of the academic year and realize that they have no notes because no one has ever taught them how to take notes during class." Another interviewee added, "My easy way out might just be, don't allow anyone to create their own references; let everyone have the standard reference aid. It's no one is getting an advantage by their ability to create a better reference aid." Importantly, a third interviewee responded, "I think in terms of teachers though, I mean there's always going to be equity issues, because there's always going to be different school context, different levels of support, different quality of teaching. That's always going to be there, whatever happens. I'm not sure, well, there's no perfect way and there won't be."

A few interviewees also noted how OBEs and the use of reference aids might impact testing accommodations. As the literature noted, use of reference aids is sometimes limited or restricted by the amount of information a student can access (e.g., one A4 sheet of paper, one 3 X 5-inch note card). Interviewees responded to this by thinking through how it would affect accommodations:

> *As once interviewee noted, "It's not fair to limit students by saying, 'You can have two sides of A4.' But if you write really tiny then you get so much more information on there but then some people may have a sight disability, can't see tiny writing, so why should they have to bring in far less notes instead of somebody else just because of that?*

Another interviewee responded, "I feel like that would be solved by special consideration… something like, 'Well, if you need maybe 25% more paper.' There are ways to work around it." A third interviewee commented from a universal design perspective saying,

> *You're helping students of all abilities. It applies to every grade level, to leverage what they have to be successful in the world. And I think this fits nicely into this conversation about open book or for our approaches to teaching and approaches to learning that we want to help students use everything they have available to them to become successful in not just*

> *their education, but in life. And so, I think it's a strong advocacy piece, especially when you put the inclusion lens on it.*

Suggestions for Future Research

Throughout the interviews, IB staff suggested future research areas that would benefit the IB. One suggestion was to engage IB teachers as subject matter experts and explore which of the existing assessments might benefit from an OBE approach. Another interviewee suggested the use of piloting or prototyping to sufficiently trial the use of reference aids prior to scaling up. A third interviewee proposed a comparative study examining the difference between OBEs and CBEs. And yet another IB staff member suggested conducting empirical research with IB schools to explore the unintended consequence of OBEs.

IB Stakeholder Survey Results

IB staff, selected IBEN representatives, and programme coordinators completed a survey on their perceptions of OBEs and the use of reference aids. Information from the surveys were used to further contextualize the literature review and the results were key in developing initial guidelines for OBEs and the use of reference aids in IB's context. The following section presents the overall survey results.

Perceptions of Open Book Assessment and Reference Aids

Survey respondents were asked whether they had experience with OBEs and the use of reference aids. The majority of respondents (about 70% on both surveys) indicated that they had some experience with OBEs. Respondents were also asked to rate their receptivity of OBEs and the use of reference aids in formative classroom assessments, summative classroom assessments, and IB summative assessments.[6] On a scale of 0 (not at all receptive) to 10 (very receptive), the average receptivity rating was highest for formative classroom assessment and was lowest for IB summative assessments. Ratings did not differ between groups.

[6] Formative classroom assessments refer to informal assessments used to inform teaching and learning (which may or may not contribute to the course grade). Summative classroom assessments refer to exams administered to evaluate student learning at the end of an instructional unit or course and contribute to the student's grade. IB summative assessments cover all IB student work that is marked or moderated by IB examiners.

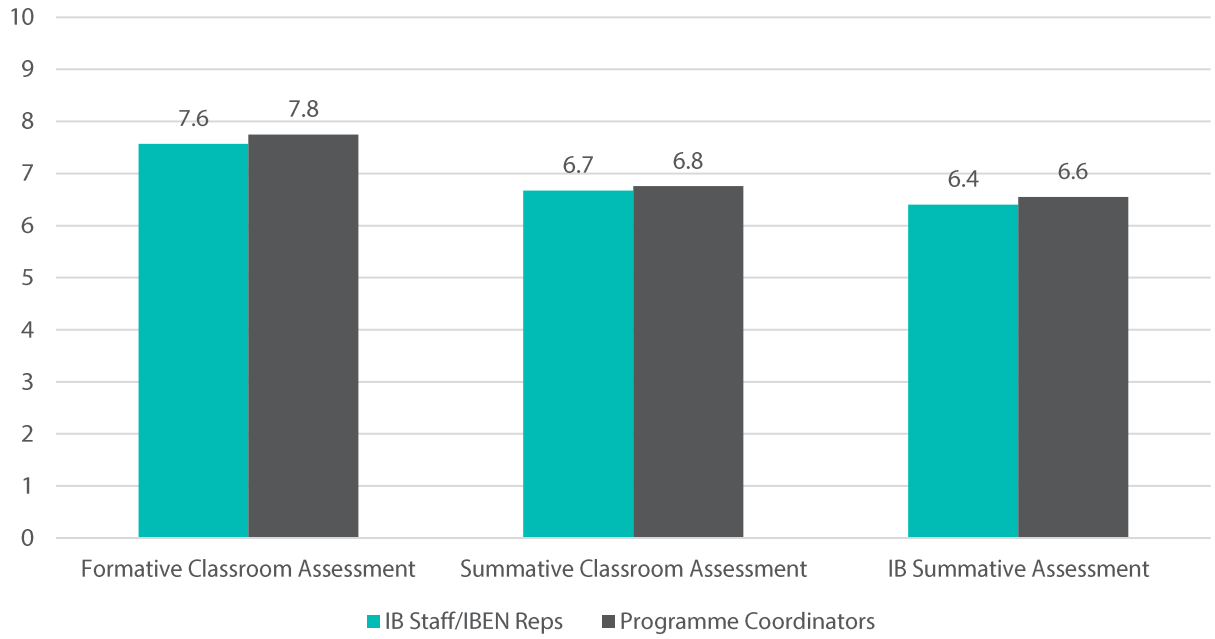

Figure 2. Receptivity of OBEs by Type of Assessment in Which OBE Would be Used.

Programme coordinators were asked to rate their perceptions of the extent to which various stakeholder groups in the school community would support or oppose OBEs on a scale of 1 (strongly oppose) to 5 (strongly support). Overall programme coordinators reported that students would be most supportive of OBEs. Although still highly rated, IB programme coordinators perceived teachers to be least supportive of OBEs.

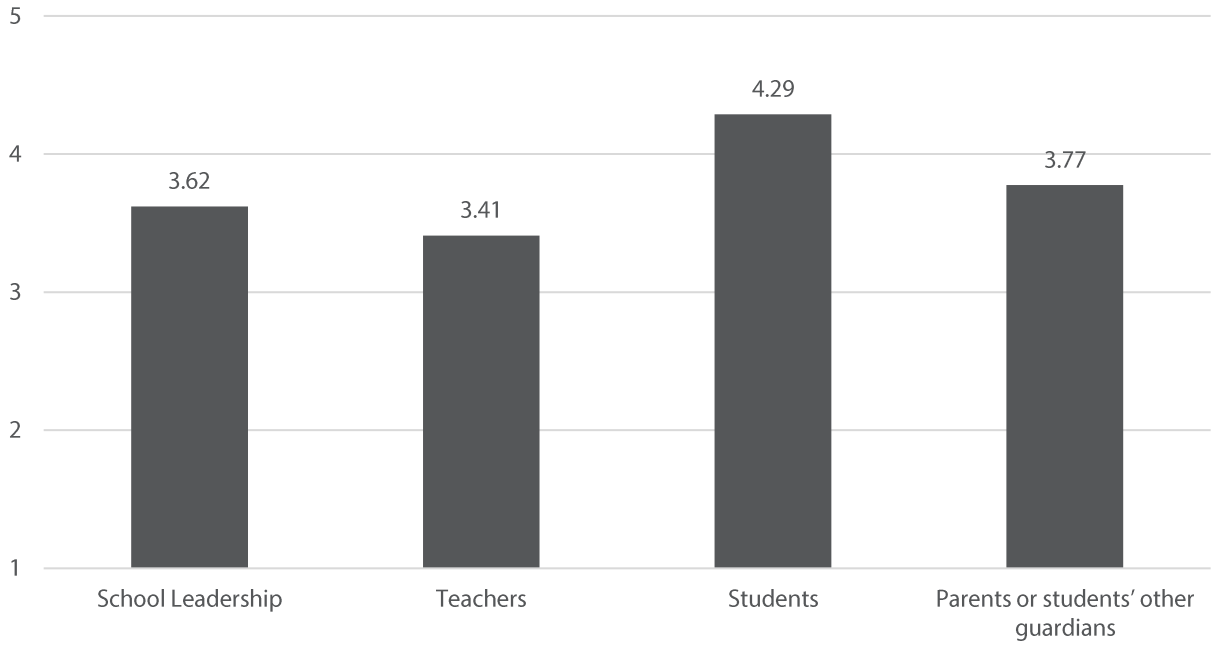

Figure 3. Perceived School Stakeholder Support of OBEs.

Key Reference Aids

Survey respondents were asked to select the reference aids that were most relevant to consider for each of the specified areas. As with all questions, respondents were permitted to only answer the parts of the question they felt confident in answering (e.g., respondents who were unfamiliar with the DP were encouraged to leave DP-specific parts of the question unanswered). Overall, programme coordinators preferred IB-prepared notes for every area except Language and Literature, and Language Acquisition. Similarly, IB staff and IBEN representatives rated IB-prepared notes and texts as the most relevant reference aids across areas. However, limited student notes were also considered relevant for Geography/History and DP Global Politics/Philosophy/Psychology/Anthropology/World Religions.

Table 6. Most Relevant Reference Aids for Different Content Areas

Key Reference Aids	IB Staff/IBEN Reps	Programme Coordinators
General Question Types:		
Lower Order Thinking Skills	Texts	IB-Prepared Notes
Higher Order Thinking Skills	Texts	IB-Prepared Notes
DP and MYP Content Areas/ Tasks:		
Language and Literature	Texts	Texts
Language Acquisition	Texts	Texts
Geography/History	Limited Student Notes	IB-Prepared Notes
Sciences	IB-Prepared Notes	IB-Prepared Notes
Mathematics	IB-Prepared Notes	IB-Prepared Notes
DP-Specific Content Areas with Exams:		
DP Business Management/Economics	IB-Prepared Notes	IB-Prepared Notes
DP Global Politics/ Philosophy/ Psychology/ Anthropology/ World Religions	Limited Student Notes	IB-Prepared Notes
DP Computer Science	IB-Prepared Notes	IB-Prepared Notes
MYP-Specific Content Areas with Exams:		
MYP Integrated Humanities	IB-Prepared Notes	IB-Prepared Notes
MYP Interdisciplinary	IB-Prepared Notes	IB-Prepared Notes

Next, survey respondents were asked to rate how easily they believed each reference aid could be incorporated into future IB summative assessments. Reference aids were rated on a 5-point scale from very difficult (1) to very easy (5). Overall, both groups agreed that IB-prepared notes or reference aids could be incorporated most easily, followed by texts (textbooks, dictionaries, literary works). Additionally, both groups agreed that restricted and unrestricted access to web resources would be

the most difficult to incorporate into IB summative assessments. Moreover, programme coordinator ratings were consistently higher than IB staff and IBEN representatives.

Table 7. Ease of Incorporation of Reference Aids

Reference Aids	IB Staff/ IBEN Reps			Programme Coordinators		
	N	*M*	*SD*	*N*	*M*	*SD*
Texts (textbooks, dictionaries, literary works)	316	3.72	1.15	276	3.75	1.14
IB-prepared notes or reference aids	311	3.79	1.05	275	4.15	0.96
Teacher prepared notes or reference aids	313	2.86	1.30	273	3.32	1.21
Limited access to student prepared notes or reference aids	317	3.04	1.25	275	3.27	1.18
Unlimited access to student prepared notes or reference aids	312	2.74	1.37	275	3.22	1.36
Restricted access to web resources	316	2.66	1.18	276	2.59	1.18
Unrestricted access to web resources	316	2.55	1.45	276	3.07	1.46

Open Book Assessment and Reference Aid Best Practices

As previously discussed, the literature suggests that open book assessment and the use of reference aids can have a variety of impacts. To explore this further, survey respondents were asked to rate the extent to which they believed the introduction of open book assessment and reference aids into IB assessments would have a negative or positive impact on key areas identified by the literature, on a scale of 0 (negative) to 10 (positive). Overall, ratings were slightly positive across areas and no significant differences were observed between groups. Generally, respondents thought OBEs would have a neutral impact on exam implementation, academic integrity, and IB's ability to assess student knowledge. Respondents generally reported OBEs would positively impact student well-being, IB's ability to assess higher order thinking skills, and student confidence in their academic ability.

Table 8. Perceptions of the Impact of Reference Aids

Open Book Assessment and Reference Aid Best Practices	IB Staff/ IBEN Reps			Programme Coordinators		
	N	*M*	*SD*	*N*	*M*	*SD*
Academic integrity	321	5.15	2.66	271	5.32	2.73
Authenticity of the assessment	320	5.69	2.68	268	5.56	2.69
Classroom assessment practices used by teachers	315	6.49	2.38	274	6.55	2.42
Curriculum and content coverage	318	6.50	2.28	274	6.74	2.30

Open Book Assessment and Reference Aid Best Practices	IB Staff/ IBEN Reps			Programme Coordinators		
	N	*M*	*SD*	*N*	*M*	*SD*
Exam and item design	317	5.97	2.71	272	6.35	2.57
Exam implementation	314	5.12	2.58	271	5.56	2.61
IB's ability to assess higher order thinking skills	318	6.91	2.66	270	6.81	2.79
IB's ability to assess student knowledge	316	5.16	2.68	267	5.42	2.77
IB's ability to focus on the skills that the task was designed to assess	314	6.56	2.67	273	6.54	2.59
IB's ability to exclude assessing skills that the task was not designed to target	311	5.72	2.61	271	5.84	2.45
Marking and grading of the exams	315	5.26	2.52	272	5.61	2.64
Pedagogical strategies used by teachers	314	6.58	2.31	271	6.60	2.43
Student confidence in their academic ability	320	6.58	2.44	272	6.71	2.59
Student performance	320	6.17	2.21	271	6.52	2.28
Student test preparation	318	5.76	2.61	266	5.84	2.71
Student well-being (stress, anxiety)	317	7.01	2.30	274	7.15	2.26
Test security	316	4.90	2.64	269	5.72	2.78

There are a number of recommendations in the literature about how reference aids should be used in assessment. To understand how these relate to the IB context, survey respondents were asked to rate the relevance of the recommendations for IB programmes on a scale of 1 (not at all relevant) to 5 (extremely relevant). Overall, respondents agreed the recommendations from the literature were relevant for IB programmes. The recommendation that was rated as most relevant for IB was to provide explicit expectation on how students should prepare for the open book exam. The lowest rated recommendation focused on using reference aids with formative assessments to facilitate studying but administering a closed book summative exam. Ratings were consistently higher for programme coordinators than for IB staff and IBEN representatives and these differences were significant for 6 of the 9 recommendations.

Table 9. Perceived Relevance of Recommendations from the Literature

Recommendations	IB Staff/ IBEN Reps			Programme Coordinators		
	N	*M*	*SD*	*N*	*M*	*SD*
Minimize the number of questions requiring factual recall on open book exams*	311	3.99	1.10	263	4.07	0.94
Offer both closed book (including recall of foundational knowledge) and open book summative exams assessing different types of learning*	311	3.96	1.09	262	3.94	0.94
Provide explicit expectation on how students should prepare for the open book exam.	310	4.53	0.85	262	4.52	0.73
Provide explicit expectation on the types and anticipated difficulty of questions that will be asked in open book exams*	311	4.33	0.98	260	4.48	0.74
Provide instruction focused on creating quality reference aids*	310	4.26	0.97	262	4.39	0.82
Provide students the opportunity to create and practice with their own reference aids*	308	4.18	1.00	263	4.37	0.83
Provide students the opportunity to practice using reference aids before the summative assessments*	308	4.37	0.90	262	4.48	0.78
Structure the exam in a way to discourage students from spending too much time searching for answers (e.g., strict time limits, higher order question)	311	4.28	0.98	261	4.33	0.96
Use reference aids with formative assessments to facilitate studying, but administer a closed book summative exam	307	3.27	1.35	258	3.46	1.26

** indicates significant difference ($p<0.05$).*

It is anticipated that schools may anticipate a variety of challenges related to implementing open book assessment and the use of reference aids. As such, survey respondents were asked to rate their level of concern that each of the provided issues may be a challenge. Statements were rated on scale of 1 (not at all concerned) to 5 (extremely concerned). Overall, respondents were only somewhat concerned with the specified challenges. Both groups were most concerned with the logistical problems of ensuring that only permitted reference aids are accessible by students. Additionally, IB staff and IBEN representatives were significantly more concerned that OBEs would lead to increased examination dishonesty during the exam than programme coordinators. Further, IB staff and IBEN representatives were least concerned with the perception that OBEs are less rigorous, while

programme coordinators were least concerned about students' and teachers' lack of familiarity or confidence with OBEs and reference aids.

Table 10. Challenges Related to OBEs

Challenges	IB Staff/ IBEN Reps			Programme Coordinators		
	N	*M*	*SD*	*N*	*M*	*SD*
Increased examination dishonesty during exams (e.g., plagiarism, cheating)*	309	3.22	1.27	262	3.16	1.18
Increased examination dishonesty prior to exams (e.g., hiding unapproved information within an approved reference aid)	309	3.31	1.25	262	3.12	1.30
Logistical problems of ensuring that only permitted reference aids are accessible by students.	307	3.57	1.15	262	3.36	1.22
Students' lack of familiarity or confidence with open book exams and reference aids	308	2.97	1.23	260	3.02	1.24
Teachers' lack of familiarity or confidence with open book exams and reference aids	307	3.19	1.21	262	3.03	1.17
The perception that open book exams are **less rigorous**	309	2.91	1.38	261	3.16	1.36
The perception that open book exams do not require students to **prepare in advance**.	310	3.10	1.41	260	3.24	1.37

** indicates significant difference ($p<0.05$).*

Assessment Design Skills

Survey respondents were asked to rate the extent to which they believed that certain assessment-related skills were needed to successfully implement reference aids in future IB summative OBEs. The scale ranged from very unimportant (1) to very important (6) The survey contained slightly different skills for IB staff and IBEN representatives and programme coordinators. Overall, all skills were rated as somewhat to very important demonstrating how critical these skills will be for successfully implementing OBEs.

Table 11. Importance of Assessment Design Skills

Assessment Design Skills	*N*	*M*	*SD*
IB Staff and IBEN Representatives			
Balancing question types to make optimal use of reference aids	301	5.22	1.12
Creating assessment tasks using reference aids	300	5.02	1.24
Designing an open book exam that includes a realistic number of items, given the time limit of the exam	303	5.24	1.11
Determining which reference aids are best suited for different curriculum areas	299	5.19	1.28
Determining which reference aids are best suited for exam questions and the skills they are designed to assess	300	5.30	1.13
Marking and grading exams that allow the use of reference aids	298	5.01	1.28
Creating reference aids for student use in an exam	294	4.69	1.42
Writing high-quality open book test questions to assess higher-order thinking skills	299	5.46	1.11
Programme Coordinators			
Balancing question types to make optimal use of reference aids	257	5.27	0.84
Creating learning activities for student to make their own reference aids	258	5.31	0.87
Creating practice tasks using reference aids	256	5.45	0.81
Designing an open book test that includes a realistic number of items, given the time limit of the test	257	5.52	0.79
Determining which reference aids are best suited for their curriculum areas	257	5.42	0.83
Determining which reference aids are best suited for their test questions and the skills they want to assess	258	5.41	0.84
Grading and providing feedback to students on tests that allow the use of reference aids	258	5.41	0.84
Creating reference aids for student use	258	5.12	1.04
Using data from tests that allow the use of reference aids to inform teaching	254	5.26	0.96
Writing high-quality open book test questions to assess higher-order thinking skills	257	5.60	0.77

School Resources and Assistance

It was anticipated that schools might need a variety of supports from the IB to successfully implement open book assessments and the use of reference aids. As such programme coordinators were asked to rate the level of importance for suggested supports. The scale ranged from very unimportant (1) to very important (6). Programme coordinators rated all supports as very important. This suggests schools will want extensive guidance and support from IB if OBEs are implemented.

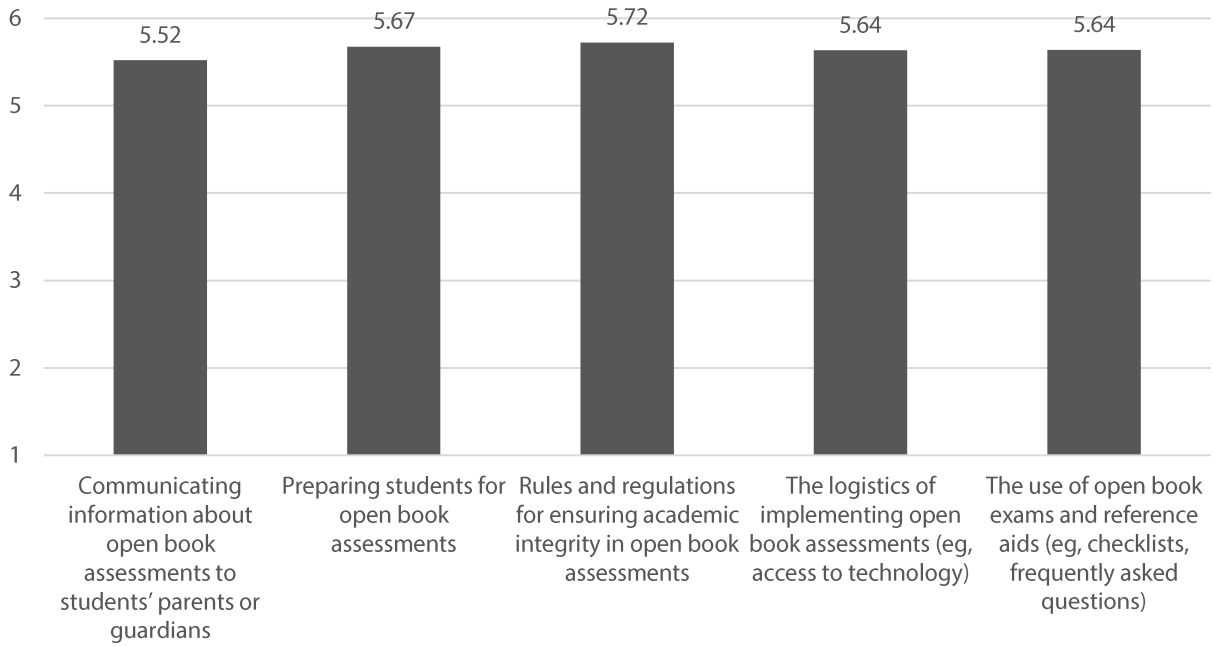

Figure 4. Importance of various supports for schools.

Programme coordinators were also asked how useful certain professional development formats would be for training school personnel on the different aspects of open book assessment and the use of reference aids. The scale ranged from 0 (not useful) to 10 (extremely useful). Again, programme coordinators rated all formats as useful.

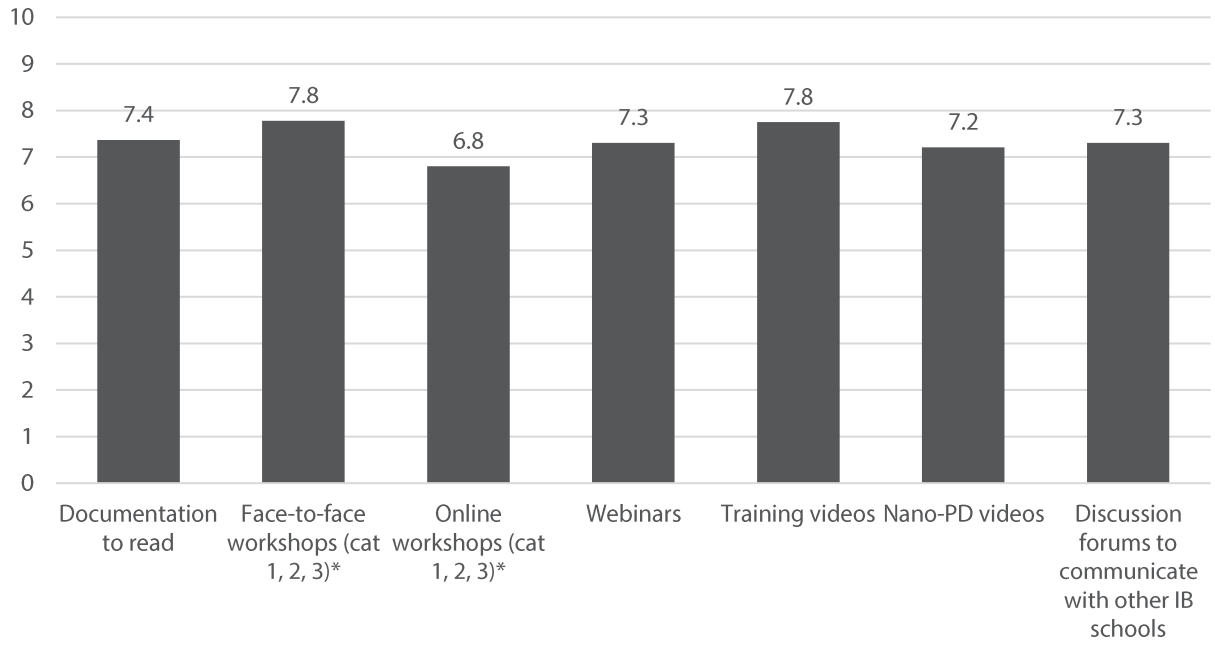

Figure 5. Usefulness of various professional development formats.

Alignment with IB Pedagogical Practices and Policies

Finally, survey respondents were asked to rate the extent to which open book practices aligned with existing IB pedagogical practices. The scale ranged from 1 (not aligned) to 5 (completely aligned). Overall, IB staff, IBEN representatives, and programme coordinators rated OBEs as being mostly aligned with IB's pedagogical practices.

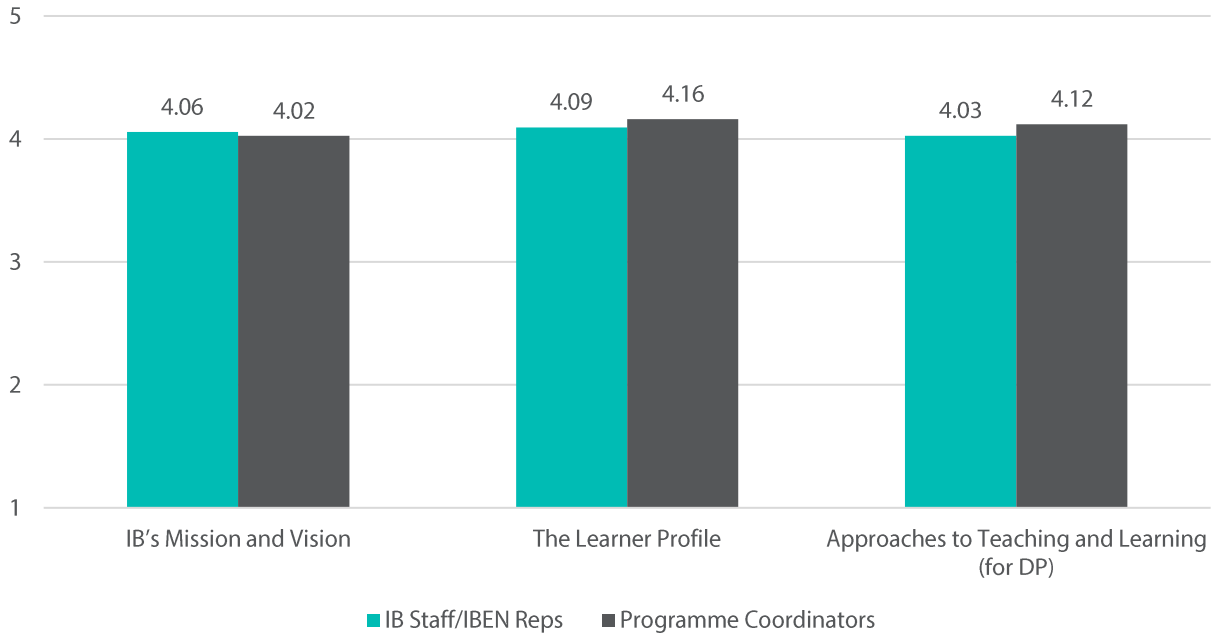

Figure 6. Perceptions of alignment of OBEs with IB pedagogical practices.

IB staff and IBEN representatives were asked to rate the extent to which open book practices aligned with existing IB policies. Again, the scale ranged from 1 (not aligned) to 5 (completely aligned). Overall, IB staff and IBEN representatives rated OBEs as being somewhat to mostly aligned with IB's policies. Respondents reported the OBEs were most aligned to policies related to access, inclusion, and universal design and were least aligned to policies related to assessment procedures.

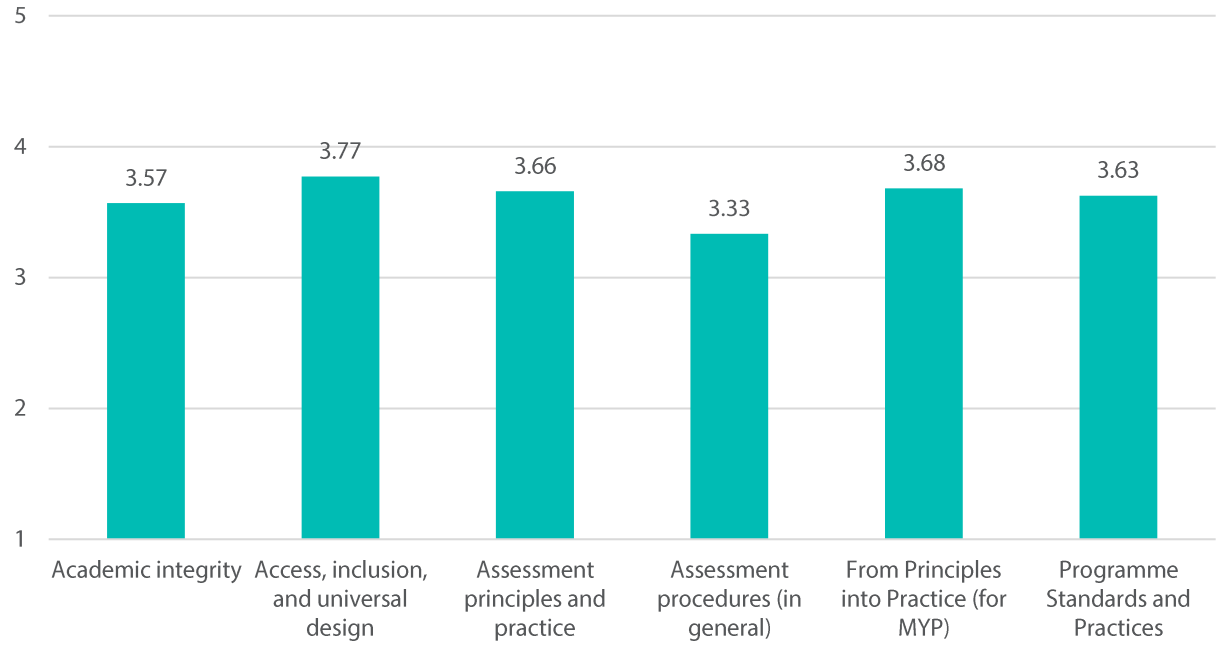

Figure 7. Perceptions of alignment of OBEs with IB policies.

Guideline Development for OBEs and Use of Reference Aids

Inflexion researchers developed initial guidelines specific to OBEs and use of reference aids on IB summative and classroom-based assessments anchored in the best available research evidence and contextualized to IB's specific context. The development of actionable guidelines was heavily dependent on the quality of information available in the literature, as well as specific decisions points for IB, such as the type of reference aids that will be used and the context around their use. However, given the fractured body of evidence and the lack of clarity around what OBEs and the use of references aids might look like in the IB, the guidelines were not as specific and actionable as originally envisioned. Thus, the following guidelines present higher-level recommendations and considerations that will need to be further articulated as IB's understanding of OBEs and decisions around their use evolve. Further, Inflexion researchers provided a subjective categorization of each guideline based on the level of importance for successful implementation of OBEs. This coding was designed to facilitate the prioritization of the guidelines and assist IB staff in deciding where to focus their resources.

Overarching and Policy-Related Guidelines

Critically Important

Guideline 1: Provide clear rules and guidelines and set clear expectations for OBEs and the use of reference aids.

- The rules should be flexible enough to apply across a range of reference aids and the range of content area exams. Alternatively, separate rules could be created for different reference aids or content areas.
- Be clear to students on what materials students can bring into the exam; what is expected of them (e.g., guidance around preparing for the exam, level of difficulty to expect, length of time available for the testing session); and appropriate use of these materials (e.g., citations for direct quotes).
- Consider providing provide additional clarity for formative or classroom OBEs, as appropriate.
- Review policies on identifying suspected academic malpractice to ensure that they align with the types of malpractice often experienced in OBEs.
- To assist with issues of academic malpractice, consider developing a comprehensive code of conduct for OBEs and having students sign an honesty pledge prior to completing the exam.

Other considerations: Any critical and cross-cutting guidelines should be considered for a policy recommendation.
While it is generally accepted that some reference aids are better suited for certain subject areas, more information is needed to inform these decisions. IB should consider a study with subject matter experts to explore the appropriateness of specific reference aids of interest for content areas that may be interested in implementing OBEs.
Depending on the decisions that IB makes regarding OBEs, teachers will need additional clarity in implementing OBEs in their classroom assessments. However, the specific information needed depends on the types of reference aids that IB implements.
IB should explore the kinds of academic malpractice that may be specific to OBE. The literature focuses strongly on plagiarism; however, depending on the type of reference aids (e.g., provided by IB, student created, web based) and the extent to which the reference aids are restricted, other forms of academic misconduct may be relevant.

Links to Related Guidelines: Guidelines 4, 5, and 6, Guideline 9, and Guidelines 10 and 11.

Critically Important

Guideline 2: Provide professional learning opportunities to all stakeholders.

- Provide professional learning opportunities internally to IB staff and externally to exam content developers, exam invigilators, markers and graders, schools, and teachers.

    - Consider having professional learning communities (PLCs) of IB staff. PLCs could plan for the implementation of OBEs and prototype or pilot aspects of OBEs and then discuss their learnings and reflections.
    - Provide an outlet, such as a Community of Practice (CoP), to allow exam content developers to learn collaboratively and share best practices and reflect on lessons learned, as appropriate.
- Provide professional learning opportunities focused on increasing knowledge of OBEs and reference aids more generally, as well as training opportunities on specific reference aids.
- Identify examples of organizations or education models that have been successful at implementing OBEs and cultivate a repository of resources and lessons learned.
- Provide high-quality examples of the types of and templates for reference aids that students may use during an OBE and the appropriateness of the different types of reference aids.
- Provide high-quality examples of and templates for well-written OBE items.
- If reference aids are text-based, provide a curated list of materials from which exam content developers choose.

Other considerations: All stakeholders will need some training around OBEs and the use of reference aids. The level of training is dependent on how involved they will be in the development and implementation of OBEs and which reference aids are being used.
Successful implementation of OBEs will require high-quality examples of items and reference aids. To assist with this, IB could learn from organizations that have already implemented OBEs successfully. IB may consider an environmental scan to identify these organizations and determine the best way to learn from their experiences. The organization search conducted as part of this study could serve as a starting point for this work. Further, though exemplars exist, IB will need to consider how to tailor these examples to its context.

Links to Related Guidelines: Guidelines 4 and 6, Guidelines 7 and 8, Guideline 10, and Guidelines 13, 14, and 15.

Critically Important

Guideline 3: Review and update policies around accessibility to cover the use of reference aids.

- All students need to have access to reference aids of equal quality, regardless of format.
- If the reference aid is web-based, ensure that students have access to the allowed reference aids (e.g., some countries block certain web content, school firewalls can make it difficult to access information, not all websites are available in all languages).
- Ensure that reference aids are equivalent across languages and free from culture biases.
- Consider how the use of reference aids will affect testing accommodations (e.g., length of time, font size for restricted reference aids).
- To address issues of equity, consider writing items that are as self-contained as possible to ensure equal access to reference aids.

Other considerations: The literature provides very little information related to accessibility and accommodations. IB will need to consider how the use of reference aids may affect policies and practices related to accessibility and accommodations.

Links to Related Guidelines: Guideline 5, Guideline 9, Guidelines 11 and 12.

Curriculum Development Guidelines

Critically Important

Guideline 4: Allocate time to teach the skill of creating and/or using reference aids.

- Creating and using reference aids by students is a teachable skill. Thus, if a course includes OBEs or reference aids, it has implications for what can go into the course because time will need to be reserved to develop the skill of creating and/or using reference aids.

Other considerations: IB should consider how to scaffold the teaching of this skill, how to leverage learning across the different reference aids and subject areas, and how to balance teaching this skill with other knowledge and skills that are part of the course.

Links to Related Guidelines: Guidelines 1 and 2, Guideline 6, and Guideline 14.

Moderately Important

Guideline 5: Determine whether use of reference aids should be a skill targeted by the assessment.

- Consider whether creating and/or using reference aids (or more broadly, evaluating information for relevance and using available information) should be included as an essential skill in the curriculum guide.

*Other considerations: Creation and/or use of reference aids will, to some extent, affect student performance; however, these skills are not typically a focus of most assessments. IB should consider how to minimize this impact, if this is a skill that IB does **not** wish to assess. However, if IB wants to assess this skill as part of the assessment, IB should consider how to best capture this skill and incorporate it into the scoring model.*

Links to Related Guidelines: Guidelines 1 and 3, Guidelines 10, 11, and 12.

Moderately Important

Guideline 6: Provide guidance for teachers on preparing students for OBEs.

- Teachers will need to incorporate opportunities to practice with OBE and reference aids into the curriculum or through practice tests or questions (regardless of whether they are created by the IB, teachers, or students).
- If reference aids are student created, teachers will need to provide instruction on how to create a high-quality reference aid.
- Teachers will need to provide formative feedback to students on creating and using reference aids.

Other considerations: The specific guidelines needed around curriculum development will depend on the type of reference aids that will be used.
More information is needed on what constitute a quality reference aid and the instructional practices to best teach the skill, as well as how to best provide feedback to students to scaffold learning.

Links to Related Guidelines: Guidelines 1 and 2, Guideline 4, and Guidelines 13 and 14.

Assessment Design Guidelines

Moderately Important

Guideline 7: Use reference aids when they are the best choice for assessing the knowledge and skills of interest

- As is best practice, and as IB already does, items and exams should be designed with the learning outcomes in mind for OBEs.
- Assessment items that incorporate reference aids should be used when they are the best choice for assessing the knowledge and skills of interest, not just because it is novel or different.
- Guidance must be provided on the types of knowledge and skills that are best suited for assessment using OBEs.

Other considerations: For subject areas interested in piloting OBEs, IB should consider examining the student learning objectives to determine whether the knowledge and skills targeted in the course lend themselves to assessment through OBEs.

Links to Related Guidelines: Guideline 2, Guideline 8, and Guideline 13.

Critically Important

Guideline 8: Leverage reference aids to assess higher order thinking skills.

- Generally, questions that allow the use of reference aids are best suited for assessing higher-level thinking and are not well suited for assessing facts and definitions.

- Exam items should be formulated in a way that students can demonstrate that they have studied and understood the material.
- Students should be asked to apply what they have learned to answer the questions.
- Examples of these types of questions include the following:
    - Focus on questions that assess comprehension, application, analysis, evaluation, interpretation, creation, and synthesis.
    - Formulate problem-based scenarios or real-world cases, requiring students to apply their skills and knowledge to the given problem or scenario.
    - Prompt students to develop their own examples to explain concepts or to come up with their own arguments.
    - Ask students to apply a concept or theory to an unfamiliar situation and problem.
    - Ask students to choose a statement of solution and defend their selection.
    - Present relevant qualitative or quantitative data and then ask students interpretative and application questions (e.g., What do the data show? What relevance do these data or does the scenario have in terms of [component of current topic]? What factors could potentially affect these data? How would you test for these?)

Other considerations: IB summative assessment already emphasize higher order thinking. IB could consider where assessment of these skills could be facilitated and enhanced with the use of reference aids.

Links to Related Guidelines: Guideline 2, Guideline 7, and Guideline 13.

Moderately Important

Guideline 9: Ensure that time constraints are realistic given the number of questions and number of reference aids

- Time management is vital to successful OBEs. Open book questions typically take longer to answer, so make sure the time allocated is realistic.
- Balance the number of questions and time it takes to ensure that students have time to use the reference aids.
- Given that fewer items can be included, balance the coverage of content across the items.
- Limit the number of different reference aids that are being used on an exam to keep it manageable for proctors and students.
- Devise clear and unambiguous questions to limit student confusion and time spent interpreting the question so students can spend their time making use of their textbook or reference aid to effectively answer the questions.

Other considerations: If the reference aid is created by IB, decide how specific the instructions should be and what information is being provided to students about the reference aid. Do students get guidance on what reference aid to use for what question? How close to the item should the reference aid appear (e.g., as part of the test packet or linked to specific questions)?

If the exam is paper-based, students will need to have access to the reference aids as part of the exam (or will need to bring hard copy references if the reference aids are teacher or student created). If the exam is digital, decide how the reference aids will be accessed by students. Will the reference aids be hard copy or digital? Will students be able to access the reference aids freely on the computer or will they need to upload the documents into a secure environment?

Links to Related Guidelines: Guidelines 1 and 3 and Guideline 12.

Assessment Implementation Guidelines

Moderately Important

Guideline 10: Provide explicit instructions for checking reference aids and exam invigilation.

- Check the appropriateness of student-created reference aids prior to students entering the secure testing environment.
- If the reference aids are digital, facilitate the upload of reference aids to the testing environment.
- If restricted web resources are allowed, sufficiently restrict access and test that all students can access the resources they should be able to access and cannot access the ones they should not be able to access.
- If two stage testing is implemented, transition from closed to open book portions of the exam.
- Observe for evidence of cheating specific to OBEs and reference aids.

Other considerations: Depending on the decisions that IB makes regarding the types of allowable reference aids and the extent to which the reference aids are restricted, this may greatly affect the school resources necessary to ensure availability of reference aids and to check the appropriateness of reference aids (e.g., copies of books, number of individuals needed to check resources, amount of time needed to check resources).

Links to Related Guidelines: Guidelines 1 and 2, Guideline 5, and Guideline 15.

Moderately Important

Guideline 11: Review and update scoring models as needed to account for OBEs and the use of reference aids.

- Most OBE items assess higher-order thinking skills. The scoring of these items emphasizes the explanation and logic of the argument over selecting the correct solution.
- Additionally, markers and graders may be stricter on things that may have been overlooked in the past (names, places, dates, exact quotes, spellings).
- Provide guidance on detecting academic misconduct specific to OBEs and use of reference aids.

Other considerations: Most OBE items are constructed response and may result in longer responses than CBE items. Markers may be able to mark fewer items/exams than they were previously. As a result, more markers may be needed to mark the same number of exams. This may also result in a larger number of markers who need training.

Links to Related Guidelines: Guidelines 1 and 3, Guideline 5, and Guideline 12.

Critically Important

Guideline 12: Ensure that exams allowing reference aids still meet adequate assessment quality standards.

- Students generally take longer to complete OBEs, which creates a tradeoff between breadth and depth and ultimately leads to fewer exam items relative to CBE. This can impact content coverage, reliability, and potentially better discrimination between higher and lower ability students.
- Effective use of resources may be implicitly assessed on OBEs, resulting in construct irrelevant variance, if this is a skill that the exam is not intended to measure.
- Without practice, training, clear expectations, and time limits, some students will likely take too much time consulting reference aids in an attempt to identify or construct an answer; this may cause the exam score to be a poor measure of their ability.

Other considerations: While there is some information in the literature on how OBEs and the use of reference aids impact the psychometric properties of the test, IB will need explore this impact on their specific tests and in their specific contexts. IB may consider the process that is used to validate exams and/or conduct a series of pilot studies to explore the impact before scaling up the use of OBEs. IB may also want to consider, a priori, an acceptable psychometric benchmark that OBEs must meet.

Links to Related Guidelines: Guideline 3, Guideline 5, Guideline 9, and Guideline 11.

Professional Learning Guidelines

Critically Important

Guideline 13: Provide learning opportunities on how to write assessment questions that optimize the use of reference aids.

- Provide a breath of high-quality reference aid examples and well-written OBE items.
- Provide opportunities to practice writing OBE items and constructing reference aids and provide feedback, if possible.

Other considerations: IB should consider training videos and/or resource sheets on the a) anatomy of a high-quality reference aid, b) the anatomy of a well-written OBE item, and c) providing effective feedback to students.

Links to Related Guidelines: Guideline 2, Guideline 6, and Guidelines 7 and 8.

Moderately Important

Guideline 14: Provide learning opportunities on how to create and/or use reference aids.

- Provide teachers professional learning opportunities on how to scaffold student learning in the creation and use of reference aids, including providing feedback to students.
- Provide teachers professional learning on teaching students effective study strategies for OBEs and how to effectively use reference aids during OBEs.
- Provide training for teachers on how to incorporate reference aids into the curricula. Since OBEs focus on higher-order thinking skills, teachers may need guidance on how to adjust their instructional approach to explicitly target these skills in ways that allow students to see the connection to future assessments.

Other considerations: Again, more information is needed on what constitute a "good" reference aid and the instructional practices that facilitate teaching this skill. IB should consider how to scaffold the teaching of the skill and provide professional learning opportunities for teachers to increase their preparedness and effectiveness in this area.

Links to Related Guidelines: Guideline 2, and Guidelines 4 and 6.

Moderately Important

Guideline 15: Provide guidance and/or learning opportunities on checking reference aids and exam invigilation.

- Explicit guidance on checking the appropriateness of student-created reference aids prior to students entering the secure testing environment and cheating specific to OBEs and reference aids.
- If the reference aids are digital, provide training on how to upload the reference aids to the testing environment.
- If restricted web resources are allowed, provide training on how to sufficiently restrict access and how to test to ensure that all students can access the resources they should be able to access and cannot access the ones they should not be able to access.
- If two-stage testing is implemented, provide guidance and/or professional learning opportunities on the transition from closed to open book portions of the exam.

Other considerations: Again, the decisions that IB makes regarding the types of allowable reference aids and the extent to which the reference aids are restricted will affect the guidance and professional learning needed to invigilate exams.

Links to Related Guidelines: Guideline 2 and Guideline 10.

Guideline Development Working Session Summary

Inflexion researchers held three working sessions around OBE guideline development. The goal of the working sessions was to react to the initial guidelines, further articulate the suggested guidelines, and to add relevant guidelines or identify areas of work for which guidelines still need to be developed. This section summarizes key points unique to those conversations; however, this section does not reiterate discussions around the literature review, IB staff interview, and IB stakeholder survey findings.

Purpose of Open Book Exams

IB staff expressed interest in more clearly understanding and seeing examples of the specific knowledge and skills that could be targeted with OBEs. Of course, this is challenging given this information is limited in the literature and it gets complicated exponentially as the layers are added (e.g., assessing different levels of knowledge and skill across different content areas across different grades with different reference aids). Further, many of the key decision are still unknown such as which reference aids should be used in which subject areas and created by whom. However, the literature emphasizes that OBEs are best suited for assessing higher order thinking skills. Further, each IB subject area has student learning objectives that are driving the curriculum and the assessment. Thus, IB could best leverage OBEs to assess student learning objectives that focus on higher order thinking skills. This is not to say that higher order thinking skills can only be assessed with OBEs. For example, another IB staff member noted that higher order thinking skills may be better assessed through coursework than on an exam. Ultimately, OBEs are *one* option and IB will need to ensure that OBEs are being used when they are most appropriate for the knowledge and skills being targeted by the assessment.

Reference Aids for Different Subject Areas

One key question left unanswered by the literature review was the extent to which different reference aids were suitable for different subject areas. Intuitively, it appears that certain reference aids would be more appropriate for some subjects but not others. For example, formula sheets would not be relevant for humanities subjects, but it would be extremely relevant for mathematics and science. IB staff generally agreed the type of reference aid would need to vary by subject area. One staff member elaborated it is unlikely that a broad rule would be able to be flexible enough to apply across subject areas because of the degree to which materials would be relevant and welcomed across areas. Another IB staff member noted there probably will be some broad rules. For instance, whenever OBEs are used, students need to have time to practice with the reference aids. However, even if there are broad rules, IB would likely need more specific rules for each subject area.

Another IB staff member noted the IB is already using reference aids in mathematics and science in the form of IB-created formula and information booklets, respectively. However, IB staff felt it might be more difficult to incorporate OBEs in the languages than in other subject areas. An IB staff member elaborated that language acquisition may be harder to visualize because of the skills that are being assessed. This is important to consider because, as mentioned previously, not all knowledge and skills

are suited for OBE. However, another IB staff member provided an example of how reference aids may be used in language acquisition. In today's world, it is appropriate to have and use dictionaries in real-life situations (i.e., Googling how to say corkscrew when shopping for something). Thus, it may be acceptable for students to search for and evaluate the appropriateness of the information as a higher order thinking skill in language acquisition. This IB staff member noted OBEs are already coming up in internal IB discussions and internal review meetings.

A staff member suggested the IB should pilot OBEs with the different type of reference aids and study the impact on student groups. These types of trials would allow IB to develop guidelines that are more specific and actionable. Further, Inflexion researchers have included a recommendation to employ subject matter experts to delve into which reference aids align best with which content areas.

Teaching for Open Book Exams

If students are asked to create reference aids, they need to be provided the opportunity to learn how to create high-quality reference aids. If students are limited by their familiarity with reference aids, their scores will not be an accurate reflection of their ability. An IB staff member added that ensuring students are familiar with the reference aids is one of the most important findings from this study. However, there is some question as to the extent to which IB can ensure that students have time to practice with reference aids.

Further, the literature does not quantify how long it takes to teach students to create and/or use reference aids. Generally, if students have more experience with OBEs, less scaffolding and practice are needed. As students learn how to create and/or use reference aids over time, teachers can spend less time developing this skill. That is, it is a skill that can be developed across multiple courses and content areas and can be built over time. The amount of time necessary to teach students how to create and use reference aids also depends upon whether the same types of reference aids are being used. Some skills are transferable; however, if the reference aids are different, students must learn the nuances and affordances of that reference aid. Further, the amount of time necessary depends on the level to which students need to master the skill. If students are being asked to create reference aids, instruction will need to involve scaffolded learning to ensure that students can create the best reference they can for the summative exam. This would take more time than giving students an hour of instruction, providing feedback once, and then expecting students to create a reference aid for use on the summative exam. Thus, a general consideration is that if IB implements OBEs, teachers will need to dedicate time to teach this skill and that limits the time available for other content and activities. Further, there should be some guidance provided to teachers on the length of time they should spend or the skill level to which students need to perform. This can directly affect student performance on OBEs if some teachers are spending an hour to do this once and others are incorporating it into teaching, learning, and classroom assessments.

IB staff also noted there may be issues of fairness when students are allowed to create their own reference aids as teachers across all IB schools may not be able to teach students to create reference aids equally well. Thus, some IB staff indicated a preference for providing reference aids and maintaining control. Conversely, other IB staff felt students should be able to create their own

reference aids because they know what they are unsure about, whereas if IB provides the material, then IB would be assuming students know and do not know certain things. Moreover, it is probable the quality of the reference aid is related to student ability, although more research is needed.

IB staff agreed these concerns are not unique to OBEs. For example, one staff member elaborated that IB already experiences issues with differences in student preparation depending on the teacher, and, to some extent, student study strategies are already assessed on CBEs. Further, one way in which students prepare for IB summative exams is to create notes and summaries to facilitate studying. However, students are not currently taught how to create good summaries, and some students are better at creating summaries than others. The incorporation of OBEs could have huge backwash effects on student learning if students were taught these skills in preparation for OBEs. This could also help to eliminate some of the differences that already exist, And, although the influence of these factors cannot be controlled, they can be minimized.

Impact on Test Score Validity

To some extent, the quality of the reference aid will be indirectly assessed on an OBE, as will the skill of using the reference aid. IB staff questioned whether creating and using reference aids should be part of the learning objectives and the syllabus for the course. Some consideration should be given to whether this is a targeted skill that should be assessed or if it is construct irrelevant variance (i.e., knowledge or skill that affects scores but is not part of what is being targeted on the assessment).

If the reference aids are *created by IB*, then all students would have access to the same information; although, student use of the reference aid may vary. However, the effect of the reference aid becomes more variable for student or teacher created reference aids. If the reference aids are *student created*, then, to some extent, assessment scores *could* reflect student's ability to create a good reference aid, identify the information that should be on the reference aid, and is contingent on the guidance and examples they receive (e.g., whether students are getting feedback from their teachers on the reference aids they created). If the reference aids are *teacher created*, then it seems reasonable that some teacher-created materials may be better than others. Thus, IB should consider how these differences may affect student performance and the accurateness of student assessment scores. Further, IB staff noted CBEs also are assessing skills that are irrelevant, like memorization. Thus, these limitations do not make OBEs unviable. Instead, IB should be aware of the challenges and minimize their impact to the extent possible.

Impact on Students

Generally, students who are prepared for the exam will be familiar with the reference aid they are using. The reference aid will serve as a safety net. Students can look up information if they forget. This will help with student stress and anxiety, and students will likely spend less time searching for information. This would also be dependent on the exam, as sometimes students are looking for examples (e.g., including an example of symbolism in an essay). There may be a number of examples in their reference aid, and it may take time because students are overwhelmed with the number of

choice available to them, whereas students who are taking a CBE could only memorize one or two examples.

Impact on Invigilation

IB summative exams are currently administered in schools and schools are responsible for training the teachers who will serve as invigilators. It is IB policy that teachers cannot serve as the invigilator for their own exam, but they may serve as the invigilator for another subject area. Invigilators ensure any materials that used in the exam are permitted. This has particular significance for OBEs. If the reference aids are student created, there must be some considerations for inspecting and returning reference aids to students. Invigilators would need instructions and guidance on what is permitted. There would also be many logistical issues to consider. IB already has policies for calculators and dictionaries. These policies could be expanded to include other reference aids. Further, guidance for invigilators may not need to be in the form of a training session. IB could use the formats that are already in place to provide guidance to invigilators. Further, an IB staff member added that OBEs may increase the necessity for unannounced exam inspections. Additional brainstorming is needed on these kinds of impacts.

Impact on Academic Misconduct

Three examples of cheating were discussed: plagiarism, students purchasing reference aids, and students including unapproved information in an approved restricted reference aids (e.g., writing notes in a book). The review of the published literature identified plagiarism as the biggest concern around OBEs and the use of reference aids. Fortunately, IB already has policies around plagiarism and those could be expanded to cover issues related to OBEs, if needed.

Students may also attempt to purchase reference aids to use on their OBEs. One staff member asked how IB could control for ghost writers. There is already a problem with students buying extended essays. It will likely be a problem for reference aids as well. These are issues that do not arise in a single school or university, but when scaled up internationally, these are important considerations. One IB staff member shared it may not be problematic for IB if students buy their reference aids because the reference aid will not serve the same purpose. That is, the reference aid would not benefit the student in the same way if the student did not create it, and students who buy reference aids may be disadvantaging themselves. This also relates to test development and writing good items. There is only so much support students will get from a reference aid if they are being asked to analyze, critique, or evaluate (e.g., higher order thinking skills). If students do not create their own reference aids, they are missing out on some key skills they would get from developing the reference aids that would benefit them on the exam.

Students including unapproved information in an approved restricted reference aid is more complicated. One of the staff members elaborated that under the current assessment model, exams need to be completed in the schools and all materials must be provided by the schools. Thus, allowing students to bring their own materials represents a risk. If reference aids and use of materials are restricted, then all materials brought into the test environment will need to be checked to ensure that

only permitted materials are being accessed. Schools would have to be responsible for this because it would not be possible for the IB to verify all reference aids for all exams. Currently, if students bring reference aids into the testing environment, policies specify that it must be a clean copy. If textbooks were allowed, then students may need two copies of each book—one for marking and a clean copy for testing. This would impact school resources. There is an acceptable level of annotation in a text; however, IB would need to unpack what is *acceptable*, which would be challenging. Additionally, if students can choose a work, then schools would need to have numerous copies of each work to ensure that all students who want to choose that work are permitted to and have the text to use as a reference aid. This would have broad implications on school resources, guidance for schools, checking and verifying student reference aids, and implications for integrity and cheating.

One IB staff member noted there are schools that push the limits of what is allowed, so IB would need to have clear guidance for schools on what is permitted. Even then, school personnel would ask questions and want specific examples. Conversely, there would be some schools that do not ask questions and will not learn the limits. The more instructions and examples IB can give, the less opportunity there will be for variable interpretation by schools.

Further, this may have implications for examiners and examiner training. Currently, examiners can raise exceptions when they suspect academic misconduct. IB staff anticipate that the occurrences of suspected misconduct could increase with the introduction of OBEs. IB would want to be develop a plan to manage those occurrences to avoid the number of exceptions becoming unmanageable. However, IB staff agreed that asking examiners to review student-created reference aids would be too much work for the examiner and distract from the task they are asked to complete. As such, suspected misconduct should be flagged for the academic honesty team within IB.

These are all important considerations and ones without definite answers. IB would want to consider how issues of academic misconduct may affect policy. One IB staff member noted that policy would likely need to change, stating that the goal would be to understand what IB wants to achieve and then creating policies that fit. Inflexion researchers noted that some of the guidelines include considerations related to student-created reference aids, providing specific guidance to students and schools on what is allowed in the reference aid, and providing guidance for the administration of the exam. However, this is an area that needs additional consideration.

Impact on Examiners

The IB will want to ensure that the introduction of OBEs does not create implicit biases in examiners. For example, knowing that students have access to reference aids may cause examiners to subconsciously be stricter on errors that are not part of the scoring criteria and may have been overlooked in the past (e.g., names, places, dates, exact quotes, spellings). Currently, marking criteria are written in a way that allow these errors. However, if examiners are unconsciously biased by the allowance of reference aids, the IB may want to be more explicit in the training process. When, for example, a marker knows that a student has access to materials and they see things like incorrect names, they could think the student is just being lazy because they have this information at their fingertips so why are they not using it? It gives a generally negative perception of the quality of the

work and what the student is able to do. IB will want to ensure that guards are in place to ensure that markers do not develop biases based on the students' access to a reference aid.

An IB staff member provided another example, describing literature where examiners were confronted with a handwritten or typed text. The text was exactly the same; however, markers were stricter on the typed text because there were unconscious biases that students could have produced a better typed paper because it takes less time and students can easily move text, whereas the examiner would not expect students to do that in a handwritten response. Another IB staff member added that markers may expect a higher order accuracy because students have access to the information. These are known situations and are not unique to OBEs; however, IB will want to explicitly monitor for these biases in the early stages of OBE implementation and adjust examiner training, as needed.

Further, OBEs may take longer to mark, which may result in the need for more examiners and may also cost more to mark. IB should consider further conversation around the number of assessments being administered and the anticipated time it will take to mark those exams, given the types of questions being asked and/or the length of the responses. It is difficult to make generalized claims about how long marking will take. These are considerations the IB will need to explore, depending on the exams and the reference aids. Additional guidance will need to be created based on those trials.

Impact on Professional Learning

The implementation of OBEs will have a significant impact on professional learning. All stakeholders will need some training around OBEs and the use of reference aids. The level of training is dependent on how involved they will be in the development and implementation of OBEs and which reference aids are being used. Ultimately, IB should consider the best way leverage existing knowledge and systems to educate IB staff on OBE. One IB staff member noted that development days, the e-assessment group, and the innovation groups may be outlets for sharing knowledge. Additionally, IB could provide an outlet, such as a community of practice (CoP), to allow teachers and assessment content developers to learn collaboratively and share their best practices. IB staff noted this was an interesting point and one that would need additional consideration. IB already creates a CoP during IB workshops where exam content developers discuss marking and review student examples. However, IB tries to maintain the distance between the paper authors and the teachers, and the paper authors and the examiners, as to not inadvertently advantage anyone with knowledge of what may be on an exam. Further, if IB changes the format of the IB summative exams, there is the potential need to retrain the examining pool. It is unclear whether training could be added to the training that is currently being offered or whether new training specific to OBEs would need to be created.

Other Logistics Related to Open Book Exams

IB staff discussed several other logical issues related to OBEs, including instructions for and placement of reference aids, e-Assessment and digital reference aids, and the collection of student created reference aids.

Instructions for and Placement of Reference Aids

IB staff discussed how explicit exam instructions should be for reference aids, including whether reference aids should be referenced in specific items or whether students will be required to evaluate when a reference aid may be useful. One IB staff member commented that for certain subjects, such as mathematics, specific references to which formula to use would be too specific; however, in other subjects, like computer science, providing explicit reference to case studies could be helpful for focusing student responses. This could be dependent on the reference aid and whether the reference aid is meant to be general or tied to a specific item. In the example of the formula booklet, students could use any formula across the mathematics exam. However, if the reference aid includes data or is a case study, it may only be relevant for a subset of questions. If the reference aids are created to go with a specific question, then it may make sense to reference the specific aid that students need to use. However, if the reference aid is more general and contains facts, definitions, or formulas, then it may make more sense to leave it open and students can determine what information they need to use to answer the questions.

Another staff member added that if the instructions were too specific, the exam would no longer be authentic, adding that in real life, students will not be told where to look for information. They will have a wealth of information that they have to sift through and determine what they need. There was some question of whether this was a skill that should be assessed on the exam.

e-Assessment and Digital Reference Aids

IB staff also wondered about the overlap between e-assessments and OBEs. With both, IB will have the opportunity to rethink the kinds of questions that are being asked and the information that is being provided. If reference aids will be IB created, it may be logistically easier to implement in an e-assessment environment.

For the MYP e-assessment, the computers are locked so that students cannot navigate from the testing platform to other documents on the computer. Currently, materials are presented as static PDFs and appear as global tools at the top of the exam. Students can access the materials, but they cannot copy and paste. In the future, if reference aids are student or teacher created, then students could bring a hard copy into the digital exam. Alternatively, these reference aids could be uploaded to the secure testing environment.

Another staff member asked whether web-based exams was something the IB would consider, noting issues with access and fairness. An IB staff member responded, saying there are online testing platforms that can control the webpages that test takers can access. However, once the IB specifies a page, that page must be available across languages. Further, given the lag time between design and administration, there are issues regarding whether the reference aid will still be available when the exam is administered. IB could use caches or mock webpages for added precautions. For DP on-screen assessments, IB has explored whitelisting websites and applications on the computer (e.g., certain Word documents, Excel, PowerPoint, PDFs). However, based on this discussion, it sounds like this would not be the first route that IB would take with reference aids.

Collection of Student Created Reference Aids

One IB staff member asked whether the IB would be interested in collecting student created reference aids as part of the submitted exam. Another staff member responded that this would add complexity, noting that it may be useful for research purposes, but anything beyond that would be too much to manage. IB staff members also noted that collecting materials from schools tends to be problematic. Further, an IB staff member added that the guidance to schools has always been that materials should only be uploaded if they are needed for the assessment (e.g., uploading the image that goes with oral commentary based on the image). If the IB did not want to collect reference aids but still wanted to be able to access them if needed, schools could be required to keep the reference aids for a certain amount of time after the exam. Currently, schools are required to keep scratch paper used during the exam, so keeping student created reference aids would be similar. Of course, schools would then need to store the reference aids, which takes space. Thus, it was generally agreed that IB would not require schools to collect student-created reference aids. However, Inflexion researchers have a recommendation that IB collect a sample of student-created reference aids in the early years to examine the breadth and depth of what is being created and to use that to inform training for teachers and guidance for schools related to creating reference aids.

Similarly, another IB staff member asked if the IB would be interested in marking a student-created reference aid as part of a portfolio. Another staff member responded that asking students to submit their reference aids for marking undermines the purpose of the reference aid. If reference aids are going to be marked, then students would focus on including information that would get them the best grade, rather than including the information they believed would be most helpful on the exam.

Conclusion

We anticipated the literature review would reveal a wealth of information on OBEs, given their acceptance in practice. From this, we expected to develop actionable guidelines to inform OBE implementation in IB's context. However, the lack of comprehensive studies made this impossible. We are coming away with many more questions than answers. As a result, the guidelines are framed as considerations or questions instead of straightforward recommendations of how to implement OBEs successfully. This is largely due to the lack of information and questions that the IB will have to consider, such as the following:

- What types of reference aids should IB consider?
- What content areas are best suited for the different types of reference aids?
- How will OBEs be implemented and rolled out to schools?
- How will OBEs be scaled up?
- What do schools need to be successful with OBEs?

Ultimately, there is a lot more to learn before IB can implement OBEs on a large scale.

Conclusions and Recommendations

This study sought to inform IB's understanding of the best practices in the use of resource aids from a practical perspective (e.g., students with specific educational and assessment needs, cultural differences, academic integrity, teacher assessment literacy and expertise, and school resources). Ultimately, the goal of the study was to develop guidelines specific to OBEs and use of reference aids on IB summative and classroom-based assessments anchored in the best available research evidence and contextualized specifically to the IB. This report presents the summative findings of the study and provides recommendations for future research.

Phase 1 Conclusions

OBEs are not a new concept in education. Some of the first peer-reviewed articles date back to the 1930s and the number of studies focused on OBEs has doubled every decade since the 1980s. However, the literature base for OBEs and the use of reference aids is unsystematic, uncomprehensive, and splintered. Most of the research on OBEs takes place in higher education and lacks depth regarding the nuances of different reference aids and breadth across subject areas. Most of the research on OBEs is correlational in nature and very few studies compare the effects of OBEs and CBEs; most research on OBEs is conducted using small-scale interventions. Much of the research on OBEs does not provide sufficient detail about the nature of the reference aids examined, with many studies simply noting students have access to reference materials.

Not all areas of interest included in the research questions proved to be represented in the literature. Multiple sources discussed student academic performance, potential backwash effect, assessing higher order thinking, student well-being, and academic integrity considerations to some extent. Far fewer sources included OBE psychometric information, required assessment expertise or design skills, considerations regarding access and inclusion, cultural differences or school resources. A common theme evident in the OBE literature is that the success of OBEs is highly dependent on context. Across all the areas of research discussed below, past research suggests the interplay of several factors determines how OBEs affect students, teachers, and schools.

Student Performance: Results on student performance are mixed and highly dependent on context with regard to OBEs and CBEs. OBEs and CBE are generally compared in terms of long-term retention after the exam; however, this type of comparison may not be appropriate, as it does not align with the purpose of OBEs and does not leverage the benefits and affordances of OBEs. Providing students opportunities to practice with OBEs and allowing students to generate their own reference aids tend to have a positive effect on exam performance. Thus, IB may want to leverage these conditions if OBEs are implemented for IB summative exams.

Student Test Preparation: Research is inconclusive on whether students prepare differently for OBEs and CBEs. In the absence of clear expectations or specific training, students will likely prepare less for OBEs than CBEs; however, preparation may also vary by students' learning orientation. Ultimately, students can be effectively taught the study strategies most conducive to preparing for OBEs and how

to effectively use reference aids during OBEs. IB should provide clear expectations to ensure that students adequately prepare for OBEs. IB may also consider guidance for schools on training students on preparing for OBEs.

Student Well-being: Much of OBE research on student anxiety and stress is not theoretically grounded and relies almost exclusively on student self-report responses on surveys or in interviews. Overall, it appears some students report less anxiety and reduced stress from OBEs relative to CBEs. However, OBEs can increase anxiety and stress when students underestimate the difficulty of OBEs, are unfamiliar with the OBE format, or have little experience preparing and using reference aids. Setting realistic expectations and providing guidance is key to reducing student text anxiety and stress. However, eliminating all anxiety and stress from exam preparation may not be beneficial if it results in students preparing less for OBEs due to overreliance on a reference aid.

Backwash Effects on Learning: OBE can provide powerful backwash effects on learning. If OBEs are designed in ways that target higher order thinking skills, (1) teachers may need to adjust their instructional approach to also explicitly target these skills in ways that allow students to see the connection to future assessments, and (2) students will have to potentially adjust their study strategies and preparation tactics for exams. To this to be true, a number of factors must be addressed and accounted for including designing assessments to target higher order thinking skills, providing clear expectations around the use of reference aids, and providing teachers and students training and practice with OBEs.

Assessment Objectives/Purpose: The little research available is inconclusive with whether OBEs promote higher order thinking. Well-designed OBE questions require students to go beyond factual knowledge to demonstrate higher order thinking, problem solving, application, and analysis. Teaching, learning, and assessment format and item design must all be aligned to effectively target and measure higher order thinking on OBEs.

Exam and Item Design: OBEs can be designed and administered in ways that do not sacrifice psychometric quality. IB will need to balance the number of items and length of time it takes to complete OBEs to provide sufficient time for students to use reference aids. This may create a tradeoff between breadth and depth of content covered on the exam, as fewer exam items can be administered on OBEs compared to CBEs. Without practice, training, clear expectations, and time limits, some students will likely take too much time consulting reference aids in an attempt to identify or construct an answer, which may affect test score validity if the scores do not accurately represent students' abilities. Further, effective use of resources may be implicitly assessed on OBEs and jeopardize test score validity if this is a skill that the exam is not intended to measure.

Assessment literacy and design competency: Very little research exists on the type of assessment literacy and design skills test authors (e.g., teachers, exam content developers) need to effectively implement OBEs. Test authors and organizations should be cognizant of OBE length because students will likely be able to answer fewer OBE items that target higher order thinking than on CBEs. Test authors need training on developing and devising effective OBE items; they should think about the

curriculum areas that are best suited to an OBE and should attempt to use OBE items to target higher-order thinking skills.

Academic Integrity: Academic integrity and test security are issues regardless of whether exams are open or closed book. Accessing unpermitted materials is discussed less than plagiarism and communicating with others (which is most relevant to take-home exams and use of unrestricted web resources). Clear expectations, honor code agreements, and cheating detection mechanisms (e.g., plagiarism detection software) can be used to combat academic dishonesty on OBEs. Further, designing items that require the application of higher order thinking skills to generate unique answers may make academic dishonesty more challenging and easier to detect.

School resources: Very little research exists on what type of school resources are needed to support and train teachers to effectively implement OBEs. Key considerations revolve around providing time and resources for students and teachers to develop reference aids, providing equal and equitable access to reference aids (e.g., printed materials, digital devices), and developing systems to potentially restrict access during OBEs (e.g., restricted web access). However, without additional information, it is difficult to say what resources schools may need.

The potential of OBE to have a positive impact on many of the areas listed above is contingent on a host of factors that might have implications for IB schools, and on internal IB processes and practices. These factors *include but are not limited to* the following:

a. Managing the perceptions of schools and students about OBEs and the use of reference aids
b. Communicating with schools about the expectations of OBEs
c. Providing training to IB staff related to OBEs and the use of reference aids
d. Providing learning opportunities to teachers related to OBEs and the use of reference aids
e. Providing guidance on opportunities for students to practice with various reference aids
f. The impact of OBEs and the use of reference aids on academic integrity and test security
g. The impact of OBEs and the use of reference aids on the DP review and inconsistences across subjects due to different curriculum review schedules
h. The impact of OBEs and the use of reference aids on assessment design and the psychometric properties of the assessment.

Each of these factors separately contributes to the success and positive backwash of OBE in the IB programmes. Attending to these factors, in tandem, unlocks an additional benefit of transitioning to OBEs: improved overall system coherency.

Phase 2 Conclusions

From a policy, standards, and practices standpoint, it appears very little needs to be changed for OBEs to fit within IB's programmatic system. One of the first steps in planning for the implementation of OBEs is to determine what reference aids are most appropriate for specific IB exams and identifying what tradeoffs may exist. Implications for access and inclusion on OBEs will be key and must include considerations around access to different reference aids, as well as geographic and socioeconomic

considerations, given that IB is active in 158 countries. As with any change to assessment, IB will need to examine the effects of the transition to OBEs on assessment quality, specifically in terms of reliability and validity. IB will need to provide documentation that clearly describes expectations to all stakeholders. Teachers will need professional development to support the effective implementation of reference aids in IB exams.

Receptivity: Generally, receptiveness of OBEs and the use of reference aids is high. Most interviewees and survey respondents reported some experience with OBEs and the use of reference aids through their own education experiences and respondents were open to the use of reference aids in formative and summative classroom assessment and IB summative testing. However, the purpose of the assessment is key. Interviewees and survey respondents recognize that reference aids should only be used if it is the best way to assess the knowledge and skills that are being targeted by the assessment.

Alignment with IB Philosophies: The use of reference aids aligns with existing IB philosophies. Most interviewees and survey respondents generally perceived OBEs and the use of references aids to be aligned to IB policies and pedagogical practices. Further, IB is already using reference aids in some content areas and has some existing policies around specific reference aids.

Ease of Incorporation of OBEs: Reference aids should be fairly easy to incorporate into IB's practice. Although it may take work, interviewees did not think it would be challenging to incorporate OBEs into IB's assessments. However, these perceptions are largely based on the alignment between OBE and IB philosophies. There are too many unknowns at this point to fully understand what the implementation of OBEs would look like for IB. Survey respondents noted it would be fairly easy to incorporate most reference aids into classroom assessments, and to a lesser extent, IB summative assessments.

Suitability of Reference Aids for Different Subjects: Some reference aids may be better suited for certain subject areas. Generally, interviewees believed specific reference aids were better suited for certain subject areas. However, it was difficult for interviewees to identify *which* reference aids were best suited for what subject areas. Interviewees focused on the reference aid or the content area separately, but it appeared there were too many unknowns for them to provide many specific examples of reference aids *for* specific content areas. Survey respondents generally preferred the use of texts, IB-created reference aids, and limited student notes compared to the use of teacher-created reference aids, unlimited student notes, restricted web access, and unrestricted web access.

Perceived Impact of OBEs: Introducing OBEs would generally have positive effects. Interviewees noted key positive effects of OBEs including authenticity of the assessment, ability to assess higher order thinking instead of recall and memorization, and the potential impact on student well-being. Survey respondents generally perceived that OBEs would have a positive (or neutral) impact on key aspects of teaching, learning, and assessment.

Relevance of Recommendations from the Literature: The recommendations from the literature are relevant for IB's work. Even with the fractured literature base, the recommendations from research

seem to be relevant for IB's work. Interviewees generally agreed that specific recommendations were relevant. Further, survey respondents consistently rated the recommendations as relevant.

Concerns Regarding OBEs: There are some concerns about OBEs. Interviewees expressed concerns about the impact OBEs could have on academic honesty and equity. However, most of the other concerns from interviewees stemmed from a lack of clarity around what reference aids may be used and in what subject areas. Survey respondents were somewhat concerned about each potential concern they were asked to rate.

Assessment Skills for OBEs: Assessment-related skills will be key for the success of OBEs. There is limited information on the teaching and assessment design skills that are necessary for teachers to use reference aids effectively. Interviewees provided little information on the assessment-related skills needed for OBEs. However, survey respondents consistently rated each of the provided skills as being important.

Professional Learning: Professional learning will be needed both internally and externally. There are too many unknowns regarding how IB might proceed with OBEs and the use of reference aids to know how this might affect schools and IB work. It is difficult to know what professional development offerings will be needed. However, interviewees and survey respondents generally noted that they will need to increase their knowledge of OBEs and specific reference aids. Training sessions, templates, and examples will be invaluable.

Impact on IB Work: IB staff are unsure how OBEs will affect their work. Given there are still several decisions to be made, IB staff were unaware of how OBEs would impact their work. Many interviewees noted that introducing OBEs and reference aids would not affect their work directly; however, they thought OBEs would impact marking and exam authoring.

Guideline Development and Working Session Conclusions

We anticipated the literature review would reveal a wealth of information on OBEs, given their acceptance in practice. From this, we expected to develop actionable guidelines to inform OBE implementation in IB's context. However, the lack of comprehensive studies made this impossible. We are coming away with many more questions than answers. There is much more to learn about OBEs before IB can implement OBEs on a large scale. Additionally, there are several decisions that IB needs to make regarding OBEs before concrete, actionable guidelines are possible, such as the following:

- What types of reference aids should IB consider?
- What content areas are best suited for the different types of reference aids?
- What existing learning objectives may be appropriate for OBEs?
- Should creation and use of reference aids be explicitly assessed?
- How will OBEs be implemented and rolled out to schools?
- How will OBEs affect students with disabilities?
- How might cultural differences and response languages affect OBEs?
- How will OBEs be scaled up?

- What do schools need to be successful with OBEs?
- How might OBE practices need to change for formative versus summative assessment?
- How might OBE practices need to change for paper-and-pencil versus digital assessments?

Areas for Future Research

Given the lack of information in the literature, there are a number of future research studies that could be conducted to better understand how OBEs and the use of reference aids would work in IB's context.

One key aim of this study was to establish the state of the field regarding the effects of using reference aids in assessments compared to assessment without these aids. The literature review revealed a dearth of information making direct comparisons between open and closed book assessments and, in many cases, the studies that did compare open and closed book exams were severely limited in design, which limited the inferences that could be made from these studies. If the IB is interested in incorporating OBEs, additional empirical evidence is needed directly comparing key outcomes of interest for open and closed book exams. This information would not only benefit the IB but would contribute significantly to the literature base on open book assessments.

Similarly, information on specific types of reference aids and content areas was lacking. Developing a more nuanced understanding of the different types of reference aids and their relevance and performance in various content areas will be a vital first step in making evidenced based decisions around the implementation of reference aids. IB should consider engaging IB teachers as subject matter experts and explore the appropriateness of specific reference aids for content areas that may be benefit from implementing OBEs, as well as which of the existing assessments might benefit from an OBE approach. Once information is gathered from subject matter experts, theoretical findings can be tested empirically. Further, IB could partner with pilot schools to implement plan-do-study-act (PDSA) cycles to sufficiently trial the use of reference aids prior to scaling up.

Further, information on what schools and teachers need to be successful with OBE was limited. Building on the survey results, IB should consider a study with schools to examine teacher assessment design skills; access to, familiarity with, and use of reference aids; and school resources needed for successful implementation of OBEs. Implementing a similar design to the on-screen assessment report (Jacovidis, et al., 2017), IB could start by identifying school typologies based on key characteristics that may influence a school's ability to successfully implement OBEs. Then, IB could explore the key facilitators and barriers to OBE implementation for each school typology. IB could also explore the supports that schools report needing to ensure successful OBE implementation by typology. This study could include the guidance that schools and teacher would like to have regarding implementing OBEs in the classroom, and the format in which schools would prefer to receive the guidance and training.

As noted previously, successful implementation of OBEs will require high-quality examples of items and reference aids. To assist with this, IB could learn from organizations that have already implemented OBEs successfully. IB may consider an environmental scan to identify these

organizations and determine the best way to learn from their experiences. The organization search conducted as part of this study could serve as a starting point for this work. Further, though exemplars exist, IB will need to consider how to tailor these examples to its context.

Though there is some information in the literature on how OBEs and the use of reference aids affect the psychometric properties of the exam, IB will need explore this impact on its specific exams and in its specific contexts. IB may consider the process that is used to validate exams and/or conduct a series of pilot studies to explore the impact before scaling up the use of OBEs. IB may also want to consider, a priori, an acceptable psychometric benchmark that OBEs must meet.

IB was interested in how OBEs related to other areas such as paper-pencil and digital assessment, access and inclusion, testing accommodations, teacher assessment design skills, and school resources needed for successful OBE implementation. Unfortunately, these topics were not covered in in the literature. IB should consider what they would hope to learn in each of these areas and then consider how to best capture that knowledge.

Final Thoughts

Ultimately, this study sought to develop actionable guidelines that could be used as IB expanded OBEs and the use of reference aids within the organization. However, after extensive searching, it became clear that the literature did not lend itself to actionable guidelines. As such, general guidelines and considerations were provided to help IB staff frame their thinking about OBEs and to inform the next steps of the OBE journey.

References

Agarwal, P. K., Karpicke, J. D., Kang, S. H., Roediger III, H. L., & McDermott, K. B. (2008). Examining the testing effect with open-and closed-book tests. *Applied Cognitive Psychology*, *22*(7), 861-876.

Agarwal, P. K., & Roediger III, H. L. (2011). Expectancy of an open book test decreases performance on a delayed closed book test. *Memory*, *19*(8), 836–852.

Berrett, D. (2012). Harvard Cheating Scandal Points Out the Ambiguities of Collaboration. *The Chronicle of Higher Education*, *59*(3).

Block, R. M. (2012). A discussion of the effect of open book and closed book exams on student achievement in an introductory statistics course. *Primus*, *22*(3), 228-238.

Brossman, B. G., Samonte, K., Herrschaft, B., & Lipner, R. S. (2017). A comparison of open book and closed book formats for medical certification exams: A controlled study. *AERA Online Paper Repository*.

Care, E., Kim, H., Vista, A., & Anderson, K. (2018). *Education system alignment for 21st century skills: Focus on assessment*. Washington, DC: Center for Universal Education at The Brookings Institution.

Centre for Teaching and Learning. (ND). A guide for academics: Open book exams. University of Newcastle, Australia: Author. Retrieved from https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjFt4DXrbjuAhUeCTQIHd5GBgkQFjAAegQIAxAC&url=http%3A%2F%2Fwww.newcastle.edu.au%2F__data%2Fassets%2Fpdf_file%2F0006%2F268980%2FOpen-Book-Exams.pdf&usg=AOvVaw1eDCwiiRog-l_2nMRW1MTv

Centre for Teaching and Learning (2020). Designing open book exams. Concordia University, Quebec, Canada: Author. Retrieved from https://www.concordia.ca/ctl/digital-teaching/online-assessment/open-book-exams.html

Christiansen, M. A., Lambert, A. M., Nadelson, L. S., Dupree, K. M., & Kingsford, T. A. (2017). In-class versus at-home quizzes: Which is better? A flipped learning study in a two-site synchronously broadcast organic chemistry course. *Journal of Chemical Education*, *94*(2), 157–163.

College Board. (2020). Advanced Placement® 2020: What students need to know. Retrieved from https://counselors.collegeboard.org/pdf/overview-presentation-ap-exams-home.pdf

Duret, D., Christley, R., Denny, P., & Senior, A. (2018). Collaborative learning with PeerWise. *Research in Learning Technology*, *26*, 1–13.

Dickson, K. L., & Bauer, J. J. (2008). Do students learn course material during crib sheet construction? *Teaching of Psychology*, *35*(2), 117–120.

Duncan, D. G. (2007). Student performance shows slight improvement when open notes are used during information systems exams. *Journal of Information Technology Education: Research*, *6*(1), 361-370.

Durning, S. J., Dong, T., Ratcliffe, T., Schuwirth, L., Artino, A. R., Boulet, J. R., & Eva, K. (2016). Comparing open book and closed book examinations: a systematic review. *Academic Medicine*, *91*(4), 583-599.

Gharib, A., Phillips, W., & Mathew, N. (2012). Cheat Sheet or Open Book? A Comparison of the Effects of Exam Types on Performance, Retention, and Anxiety. *Psychology Research*, *2*(8), 469-478.

Gafoor, K. A., & Kurukkan, A. (2016). Self-regulated learning: A motivational approach for learning mathematics. *International Journal of Education and Psychological Research*, *5*(3), 60–65.

Grant, M. & Booth, A. (2009) A Typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal, 26,* 91-108.

Guest, G., MacQueen, K. M., & Namey, E. E. (2012). *Applied thematic analysis*. Thousand Oaks, CA: Sage.

Hamouda, S., & Shaffer, C. A. (2016). Crib sheets and exam performance in a data structures course. *Computer Science Education*, *26*(1), 1–26.

Heijne-Penninga, M., Kuks, J. B. M., Hofman, W. H. A., Muijtjens, A. M., & Cohen-Schotanus, J. (2013). Influence of PBL with open book tests on knowledge retention measured with progress tests. *Advances in Health Sciences Education*, *18*(3), 485–495.

Heijne-Penninga, M., Kuks, J. B. M., Schönrock-Adema, J., Snijders, T. A. B., & Cohen-Schotanus, J. (2008). Open book tests to complement assessment-programmes: analysis of open and closed book tests. *Advances in Health Sciences Education*, *13*(3), 263–273.

International Baccalaureate Organisation. (2013). IB learner profile. Retrieved from https://www.ibo.org/contentassets/fd82f70643ef4086b7d3f292cc214962/learner-profile-en.pdf

International Baccalaureate Organisation. (2019). Assessment principles and practices—Quality assessments in a digital age. Cardiff, Wales: Author.

International Baccalaureate Organisation. (2018). The conduct of IB Diploma Programme examinations. Cardiff, Wales: Author.

International Baccalaureate Organisation. (2017). Candidates with assessment access requirements Cardiff, Wales: Author.

International Baccalaureate Organization. (2017). *What is an IB education?* Retrieved from https://www.ibo.org/globalassets/what-is-an-ib-education-2017-en.pdf

International Baccalaureate (IB). (2021a). *About the IB*. Retrieved from http://www.ibo.org/about-the-ib/

International Baccalaureate (IB). (2021b). *Mission*. Retrieved from http://www.ibo.org/about-the-ib/mission/

International Baccalaureate (IB). (2021c). *Facts and figures*. Retrieved from http://www.ibo.org/about-the-ib/facts-and-figures/

Jacovidis, J. N., Chadwick, K. L., Fukuda, E., Todd, J., & Thier, M. (2018). *Study of International Baccalaureate Diploma Programme and Career-Related Programme school readiness for and perceptions of on-screen examinations*. Eugene, OR: Inflexion.

Jensen, P. A., & Moore, R. (2009). Students' Perceptions of Their Grades Throughout an Introductory Biology Course: Effect of Open Book Testing. *Journal of College Science Teaching*, *38*(3), 58–61.

Johanns, B., Dinkens, A., & Moore, J. (2017). A systematic review comparing open book and closed book examinations: Evaluating effects on development of critical thinking skills. *Nurse Education in Practice*, *27*, 89-94.

Karagiannopoulou, E., & Milienos, F. S. (2013). Exploring the relationship between experienced students' preference for open-and closed book examinations, approaches to learning and achievement. *Educational Research and Evaluation*, *19*(4), 271–296.

Kim, J. (2019). The Effects of Note-Taking Strategy Training on Students' Notes During Academic English Listening Tests. *English Teaching, 74*(1).

Lam, W., Williams, J. B., & Chua, A. Y. (2007). E-xams: harnessing the power of ICTs to enhance authenticity. *Journal of Educational Technology & Society*, *10*(3), 209–221.

Larwin, K. H., Gorman, J., & Larwin, D. A. (2013). Assessing the impact of testing aids on post-secondary student performance: A meta-analytic investigation. *Educational Psychology Review*, *25*(3), 429-–443.

Leaf, David E, Leo, Joseph, Smith, Phillip R, Yee, Herman, Stern, Arnold, Rosenthal, Pamela B, … Pillinger, Michael H. (2009). SOMOSAT: Utility of a web-based self-assessment tool in undergraduate medical education. *Medical Teacher*, *31*(5), e211–e219.

Lindblom-Ylänne, S., & Lonka, K. (2001). Students' perceptions of assessment practices in a traditional medical curriculum. *Advances in Health Sciences Education*, *6*(2), 121–140.

Lipner, R. S., Brossman, B. G., Samonte, K. M., & Durning, S. J. (2017). Effect of access to an electronic medical resource on performance characteristics of a certification examination: A randomized controlled trial. *Annals of Internal Medicine*, *167*(5), 302–310.

Myyry, L., & Joutsenvirta, T. (2015). Open book, open-web online examinations: Developing examination practices to support university students' learning and self-efficacy. *Active Learning in Higher Education*, *16*(2), 119–132.

Nevid, J. S., Pyun, Y. S., & Cheney, B. (2016). Retention of text material under cued and uncued recall and open and closed book conditions. *International Journal for the Scholarship of Teaching and Learning*, 10(2), 1–4.

Nsor-Ambala, R. (2020). Impact of exam type on exam scores, anxiety, and knowledge retention in a cost and management accounting course. *Accounting Education*, *29*(1), 32–56.

Phillips, G. (2006). Using Open-Book Tests to Strengthen the Study Skills of Community-College Biology Students. *Journal of Adolescent & Adult Literacy*, *49*(7), 574–582.

Ramamurthy, S., Er, H. M., Nadarajah, V. D., & Pook, P. C. (2016). Study on the impact of open and closed book formative examinations on pharmacy students' performance, perception, and learning approach. *Currents in Pharmacy Teaching and Learning*, *8*(3), 364–374.

Rice, A. H., Vogelweid, C. M., & Kitchel, T. (2017). The influence of crib sheets on veterinary students exam performance, perceived stress, and retention of subject matter knowledge. *NACTA Journal*, *61*(1), 66–72.

Stalnaker, J. M., & Stalnaker, R. C. (1934). Open book examinations. *The Journal of Higher Education*, *5*(3), 117-120.

Theophilides, C., & Koutselini, M. (2000). Study behavior in the closed book and the open book examination: A comparative analysis. *Educational Research and Evaluation*, *6*(4), 379–393.

University of Oxford. (2020). Honour code for open book and closed book remotely invigilated online exams. Retrieved from https://www.ox.ac.uk/students/academic/exams/open-book/honour-code

Vogelweid, C. M., Kitchel, T., & Rice, A. H. (2014). Veterinary students' use of crib sheets in preparing for learning and reducing stress. *NACTA Journal*, *58*(2), 135–139.

Weber, L. J., McBee, J. K., & Krebs, J. E. (1983). Take home tests: An experimental study. *Research in Higher Education*, *18*(4), 473–483.

Williams, J. B. (2006). The place of the closed book, invigilated final examination in a knowledge economy. *Educational Media International*, *43*(2), 107–119.

Williams, J. B., & Wong, A. (2009). The efficacy of final examinations: A comparative study of closed-book, invigilated exams and open-book, open-web exams. *British Journal of Educational Technology*, *40*(2), 227–236.

Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*. New York, NY: ACM.

Appendices

Appendix A: Full List of Literature Review Search Terms

Search Terms
ab("open book" OR "closed book")
ab("test* aid*")
ab ("exam* aid*")
ab("assessment* aid*")
ab(web-based OR internet-based OR paper-based) AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab("student note*") AND ab(assessment* OR test* OR exam*)
ab("crib sheet") AND ab(assessment* OR test* OR exam*)
ab(assessment* OR test* OR exam*) AND ab("clue card")
ab("memory aid") AND ab(assessment* OR test* OR exam*)
ab(dictionar*) AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab("graphic organizer") AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab("practice exam") AND ab(assessment* OR test* OR exam*)
ab("mock exam") AND ab(assessment* OR test* OR exam*)
ab("specimen exam") AND ab(assessment* OR test* OR exam*)
ab("past exam") AND ab(assessment* OR test* OR exam*)
ab("assignment") AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab("textbook*") AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab("process portfolio*") AND ab(assessment* OR test* OR exam*)
ab("graphic organiser") AND ab(assessment* OR test* OR exam*)
ab("literary text*") AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab("fact sheet*") AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab("fact sheet*") AND ab(assessment* OR test* OR exam*)
ab("literary text*") AND ab(assessment* OR test* OR exam*)
ab("book*") AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab("video*") AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab("data") AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab("image*") AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab("organiz*") AND ab(aid) AND ab(assessment* OR test* OR exam*)

Search Terms
ab("organis*") AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab("formula booklet*") AND ab(assessment* OR test* OR exam*)
ab("student summar*") AND ab(assessment* OR test* OR exam*)
ab("reference guide") AND ab(assessment* OR test* OR exam*)
ab("reference guide") AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab(computer-assist* OR computer-aid*) AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab("grammar table") AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab("graphic organizer") AND ab(assessment* OR test* OR exam*)
ab("digital portfolio") AND ab(assessment* OR test* OR exam*)
ab("open book" OR "closed book") AND ab(assessment* OR test* OR exam*)
ab(computer-assist* OR computer-aid*) AND ab(assessment* OR test* OR exam*)
ab("list of formulae") AND ab(assessment* OR test* OR exam*)
ab("statistical tables") AND ab(assessment* OR test* OR exam*)
ab("data booklet*") AND ab(assessment* OR test* OR exam*)
ab("Wolfram alpha") AND ab(assessment* OR test* OR exam*)
ab("graphing package*") AND ab(assessment* OR test* OR exam*)
ab("spreadsheet*") AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab(stress) AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab("Generalized Anxiety Disorder") AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab("mental health") AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab("depression") AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab("English Second Language Learners") AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab("English Language Learners") AND ab(aid) AND ab(assessment* OR test* OR exam*)
ab("universal design") AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab("universal design of assessment")
ab("universal design of learning")
ab("special educational needs") AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab("special needs") AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab("access arrangement*") AND ab(assessment* OR test* OR exam*)
ab("accommodation*") AND ab(assessment* OR test* OR exam*) AND ab(aid)

Search Terms
ab("additional time*") AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab("extra time*") AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab("reader") AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab("reader software") AND ab(assessment* OR test* OR exam*)
ab("reader aloud") AND ab(assessment* OR test* OR exam*)
ab("braille") AND ab(assessment* OR test* OR exam*)
ab("braille") AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab("vision impairment") AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab("low vision") AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab("visual processing challenge*") AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab(dyslexia) AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab(dysgraphia) AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab(dyspraxia) AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab(Dyscalculia) AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab("specific learning disabilit*") AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab("specific learning difficult*") AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab("processing difficult*") AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab("executive processing challenge*") AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab("working memory difficult*") AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab("short term memory challeng*") AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab(mnemonics) AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab("reading fluency) AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab(speed) AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab(accuracy) AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab(comprehension difficult*) AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab("organization difficult*") AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab("language learner*") AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab("language disorder*") AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab("communication disorder*) AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab(autism) AND ab(assessment* OR test* OR exam*) AND ab(aid)

Search Terms
ab("Pervasive Developmental Disorder*") AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab(Aspergers) AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab("Attention deficit hyperactivity disorder") AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab("cerebral palsy") AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab("fine motor challeng*") AND ab(assessment* OR test* OR exam*) AND ab(aid)
ab("open source exam*)
ab("two-stage exam*")

Appendix B: Full List of Organizations and Geographic Regions for Targeted Search

Organizations
eAssessment Association
AACTE
Achieve
American Bar Association
American Board of Internal Medicine
Asia-Pacific Educational Assessment Conference
Assessment Network
Association for Educational Assessment (Europe)
Association for the Assessment of Learning in Higher Education
Australia (NAPLAN)
Brazil
CAEP
Center for Assessment: NCIEA
Centre for Teaching and Learning (University of Newcastle) newcastle.edu.au/
Chartered Insurance Institute
College Board
CSAI: The Center on Standards & Assessment Implementation
ETS
Federal Aviation Administration
Finland
Hong Kong Examination and Assessment Authorities
India Union Public Service Commission
INSPERA
International Association for Education Assessment
Israel National Institute for Testing & Evaluation
Japan
Mexico (CENEVAL)

Organizations
National Board of Medical Examiners
National Council of Examiners for Engineering and Surveying
NCATE
NCEE
NCME
NEBOSH
New Meridian
NWEA
OCDE
Ofqual
Pearson
PIRLS/TIMMS
Smarter Balanced
South Africa (UMALUSI)
South Korea
TEAC
Turkey Council of Higher Education (YÖK)
University of Bristol
University of Oxford
WestEd

Appendix C: Full List Documents for IB Document Review

IB Documents for Review

- Policy
  - Academic Integrity
  - IB assessment principles and practices: A guide to assessment for chief examiners and principle examiners
  - IB assessment principles and practices: A guide to assessment for examiners
  - IB assessment principles and practices: A guide to assessment for teachers and coordinators
  - IB assessment principles and practices: A guide to assessment for students and their parents/guardians
  - Assessment principles and practices—Quality assessments in a digital age
  - The responsibilities of IB World Schools for uploaded and submitted work
  - Academic integrity for internal assessments
  - Academic honesty in the IB educational context
  - Middle Years Programme Assessment procedures 2021
  - DP Policy Ecosystem

- Procedural
  - Diploma Programme Assessment procedures 2021
  - Conduct of examinations: Items not permitted (DP/CP)
  - Conduct of examinations: Items not permitted (MYP)
  - The conduct of IB Middle Years Programme on-screen examinations
  - Conduct of examinations booklet 2021
  - The conduct of IB Diploma Programme examinations
  - MYP on-screen examinations: IT requirements and school responsibilities

- Guidance
  - Assessment readiness 2019: Your guide to a successful exam session
  - On-screen exams: Information for candidates
  - Guide to the MYP exam session
  - Preparing for the MYP exam session
  - MYP on-screen examinations user guide

- Examples of permitted and non-permitted reference aids
  - MYP Mathematics formula booklet
  - Rough / Scratch Paper document for MYP on-screen exams
  - MYP Extended mathematics formula booklet