



Final Report:
Alignment and Coherence of Language Acquisition
Development in the International Baccalaureate
Middle Years Programme

Submitted March 17, 2017 by
Dr. Kristine Chadwick, Senior Director
Educational Policy Improvement Center
1700 Millrace Drive, Eugene OR 97403
kristine_chadwick@epiconline.org
541.246.2600

Prepared by
Michael Thier, MAT
Erin Fukuda, PhD
Stephanie Knight, MALAS
Julie Sykes, PhD
Kristine Chadwick, PhD

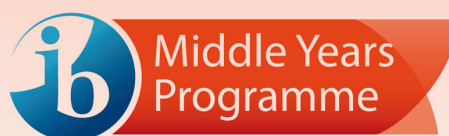


TABLE OF CONTENTS

Exectutive Summary	1
Phase I: Literature Review	1
Phase II: Within-Document Analysis	2
Phase III: Cross-Document Analysis	2
Phase IV: Progression Analysis	2
Phase V: Discrepancy Analysis (Conclusion)	3
Introduction	4
Methodological Approach	5
Phase I: Literature Review	5
<i>Data Collection</i>	5
<i>Search Parameters and Data Analysis</i>	6
<i>Validation and Report of Findings</i>	6
Phase II: Within-Document Analysis	7
<i>Recruitment and Training of Reviewers</i>	7
<i>Review Tool and Analytical Procedures</i>	7
<i>Report of Findings</i>	8
Phase III: Cross-Document Analysis	8
<i>Recruitment and Training of Reviewers</i>	8
<i>Data Sources and Analytical Procedures</i>	8
<i>Report of Findings</i>	9
Phase IV: Progression Analysis	9
<i>Validation and Report of Findings</i>	10
<i>Report of Findings</i>	10
Phase V: Discrepancy Analysis	10
<i>Data Collection</i>	10
<i>Data Analysis and Report of Findings</i>	10
Findings	12
Phase I: Literature Review	12
<i>Highlights</i>	12
<i>Recommendations Based on Literature Review</i>	13
Phase II: Within-Document Analysis	14
<i>Assessment Criteria Phase-Specific Descriptions (Clarity)</i>	15
<i>Phase-Specific Language Acquisition Objectives (Clarity)</i>	19
<i>Phase-Specific Language Acquisition Objectives (Alignment)</i>	23
<i>Global Proficiency Table (Clarity)</i>	24
<i>Global Proficiency Table (Alignment)</i>	24
<i>Language Acquisition Continuums (Clarity)</i>	25
<i>Language Acquisition Continuums (Alignment)</i>	26
<i>Phase-Specific Assessment Task Guidelines (Appropriateness)</i>	26
<i>Using Progressions to Place Students and Plan Instruction</i>	27
Phase III: Cross-Document Analysis	28
<i>Comparing Documents</i>	29
<i>Unique Strengths</i>	32
Phase IV: Progression Analysis	34
<i>Conceptualization of Effective Language Learning and Teaching</i>	37
<i>Intercultural Competence</i>	38

<i>Organizing Features of Learning</i>	38
<i>L1/L2 Integration</i>	39
<i>Phases of Language Acquisition</i>	40
<i>Language Frameworks—Areas of Communication</i>	40
<i>Language Frameworks—Communication Continuums</i>	41
<i>Language Frameworks—Communicative Processes</i>	42
<i>Approach to Assessment</i>	43
<i>Overall Progression Findings</i>	45
Phase V: Discrepancy Analysis (Conclusion)	46
Convergence	46
Complementarity	47
Contradiction	48
Next Steps	50
References	54
Appendix A: Annotated Bibliography	56
I. Phases of Language Acquisition	56
II. Communicative Processes and Language Continuums	64
III. Assessment	74
Appendix B: Full-Length MYP Literature Review	79
Purpose	79
Overview	79
Method	79
Phases of Language Acquisition	81
<i>Defining Phases of Language Acquisition</i>	81
<i>Limitations of Phases in Language Acquisition: Empirical Evidence</i>	83
<i>Limitations of Phases in Language Acquisition: Complex Progression</i>	83
<i>Limitations of Phases in Language Acquisition: Proliferation</i>	83
<i>Further Complications: Research Considerations</i>	84
Four Language Processes and the Blending of Continuums	84
Effective Input, Comprehension and Interaction, and Output	86
<i>Input</i>	86
<i>Comprehension and Interaction</i>	86
<i>Output</i>	87
Blending of Language Continuums in Language Processes: A Theoretical Perspective	87
<i>Intercultural Competence</i>	87
<i>Contextualized Communication</i>	88
<i>Multimodality and Multiliteracies</i>	88
Assessment	88
<i>The Blending of Language Continuums in Assessment Criteria: A Practical Perspective</i>	89
<i>Authentic Assessment and the Limitations of Assessment Criterion A and Assessment Criterion C</i>	89
<i>Comprehensibility of Feedback for Learners: Areas of Communication in Combination</i>	90
<i>Summative and Formative Assessment Link</i>	90
Conclusion	91
Appendix C: Within-Document Review Tool	92
Appendix D: Cross-Document Review Tool	93
Appendix E: Continuum-School Teacher Interview Protocol	95
Appendix F: Head of MYP Development Interview Protocol	96

LIST OF TABLES

Table 1. <i>Reviewers' Interrater Reliability Statistics by Task</i>	15
Table 2. <i>Number of Reviewers (N = 5) Rating Phase-Specific Objectives as Aligned by Phase</i>	23
Table 3. <i>Number of Reviewers (N = 5) Rating Global Proficiency Tables as Aligned by Phase</i>	25
Table 4. <i>Number of Reviewers (N = 5) Rating Phase-Specific Objectives as Clear by Phase</i>	25
Table 5. <i>Number of Reviewers (N = 5) Rating Phase-Specific Assessment Task Guidelines as Appropriate</i>	26
Table 6. <i>Reviewers' Interrater Reliability Statistics by Document</i>	28
Table 7. <i>Comparative Ratings of the MYP Guide and Three International Documents</i>	30
Table 8. <i>Progression Analysis of Language Acquisition in the PYP, the MYP, and the DP</i>	35
Table 9. <i>Organizing Features of Learning Across the IB Continuum</i>	39
Table 10. <i>L1/L2 Integration Across the IB Continuum</i>	40
Table 11. <i>Phases of Language Acquisition Across the IB Continuum</i>	40
Table 12. <i>Areas of Communication Across the IB Continuum</i>	41
Table 13. <i>Communication Continuums Across the IB Continuum</i>	41
Table 14. <i>Communicative Processes Across the IB Continuum</i>	43
Table 15. <i>Approach to Assessment Across the IB Continuum</i>	44
Table A1. <i>Alignment of MYP Criterion C, CEFR Qualitative Aspects of Spoken Language Use, and ACTFL Performance Descriptors</i>	82
Table A2. <i>Canale & Swain's Components of Communicative Competence Described and Related to IB Assessment Criteria</i>	85

LIST OF FIGURES

<i>Figure 1.</i> Number of reviewers ($N = 5$) rating Comprehension Spoken/Visual (Criterion A) as Clear by phase and achievement level	16
<i>Figure 2.</i> Number of Reviewers ($N = 5$) rating Comprehension Written/Visual (Criterion B) as Clear by phase and achievement level	16
<i>Figure 3.</i> Number of reviewers ($N = 5$) rating Communicate in Response (Criterion C) as Clear by phase and achievement level	17
<i>Figure 4.</i> Number of reviewers ($N = 5$) rating Use Language in Spoken or Written Form (Criterion D) as Clear by phase and achievement level	17
<i>Figure 5.</i> Number of reviewers ($N = 5$) rating phase-specific language acquisition objective as Clear by phase and strand for Criterion A	20
<i>Figure 6.</i> Number of reviewers ($N = 5$) rating phase-specific language acquisition objectives as Clear by phase and strand for Criterion B	20
<i>Figure 7.</i> Number of reviewers ($N = 5$) rating phase-specific language acquisition objective as Clear by phase and strand for Criterion C	21
<i>Figure 8.</i> Number of reviewers ($N = 5$) rating phase-specific language acquisition objective as Clear by phase and strand for Criterion D	21

EXECUTIVE SUMMARY

The International Baccalaureate (IB) has authorized more than 1,300 schools in about 145 countries to offer its Middle Years Programme (MYP) to learners aged 11 to 16. One of the MYP's unique features is its focus on second-language acquisition, which the MYP codifies in its *Language Acquisition Guide* (MYP *Guide*). The Educational Policy Improvement Center (EPIC) in Eugene, OR, U.S., partnered with the University of Oregon's Center for Applied Second Language Studies (CASLS) to examine alignment and coherence of the MYP *Guide*. The project occurred between July 2016 and January 2017 and involved five phases:

1. We *reviewed research literature* on language acquisition development and practices to examine how effectively the MYP *Guide* conceptualizes and describes a progression of learning an additional language.
2. We recruited and trained IB programme development and assessment staff¹ and school-based educators to conduct a *within-document analysis*, examining how effectively the MYP *Guide*'s assessment criteria aligned with its other progressions of language learning.
3. We recruited and trained IB staff and school-based educators to conduct a *cross-document analysis*, examining the MYP *Guide* in comparison to well-known language acquisition frameworks from three international sources.
4. In our *progression analysis*, we identified points of alignment and misalignment between the MYP *Guide* and the language acquisition guides of the Primary Years Programme (PYP) and Diploma Programme (DP).
5. In a *discrepancy analysis*, we reexamined findings from the first four phases to see how they converged with, complemented, or contradicted one another. In this phase, we asked: What changes or refinements to the MYP *Guide* would promote good teaching practice and efficiency when evaluating student progress in language acquisition?

After the executive summary, this report presents (a) a detailed description of the methods that correspond to each phase, (b) itemized findings for each of the first four phases, and (c) a conclusion that presents the discrepancy analysis and four questions to guide IB's next steps.

Phase I: Literature Review

In our review, we emphasized studies for MYP-age learners and studies completed in the past 10 years, but ultimately drew upon 20 years of peer-reviewed studies that offered insights into effective progressions of second-language learning and specified listening and speaking, viewing and interpreting, reading comprehension, and writing. Using that literature pool (see Appendix A), we found various strengths and recommendations. Strengths of the MYP *Guide* include its direction that educators provide opportunities for learners to engage in higher-level and meaningful communication activities and to foster their growth in language proficiencies. The MYP *Guide* shows communication accurately as a complex, multimodal, contextual, and integrative activity. Furthermore, considerable evidence revealed the MYP's focus on intercultural competence, aligning well with IB's mission. Moreover, the MYP *Guide* aligns with international standards, though neither completely nor directly. The literature review led to three recommendations:

¹ Hereafter IB staff.

(a) clarifying the selection of language continuums, language processes, and how they interact; (b) requiring learners to engage in the mutually influencing processes of interpreting and producing language on single assessments to avoid complicating feedback loops; and (c) providing guidance to help educators conceptualize a proficiency-based classroom.

Phase II: Within-Document Analysis

Our reviewers examined the effectiveness of the *MYP Guide*, specifically rating its components for clarity, alignment, and/or appropriateness. Reviewers demonstrated consensus that the *MYP Guide's* progressions are generally effective, though each reviewer offered suggestions for revision. First, imprecise use and/or inconsistent application of terminology reduced clarity in some portions of the *MYP Guide*, yielding a need for more explicitly defined terms. Second, educators could benefit from examples that illustrate the *MYP Guide's* expectations. Third, the *MYP Guide* applies appropriateness and/or logic inconsistently. Fourth, differentiating several strands² from one another more sharply would aid clarity and improve progressions. Fifth, IB should consider the appropriateness of using mother tongue and/or language of instruction and reexamine word-length ranges in assessment tasks.

Phase III: Cross-Document Analysis

Our reviewers reexamined the *MYP Guide*, this time in comparison to one of three international standard sets on language acquisition. Generally, reviewers rated the *MYP Guide* and the *Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (CEFR) as providing better guidance than the American Council on the Teaching of Foreign Languages' (ACTFL) *Standards for Foreign Language Learning in the 21st Century* or the Hong Kong Education Bureau's *Key Learning Areas: English and Chinese Language Education* (HKEB).³ The *MYP Guide* showed strengths in interpersonal communication; defining phases of language acquisition; communicative competence; effective input, comprehension and interaction, and output; intercultural competence; and contextualized communication. The *MYP Guide* showed room for improvement in assessment, multimodality, and multiliteracies. Findings also revealed opportunities for the *MYP Guide* to align further with the international standards.

Phase IV: Progression Analysis

Drawing on findings from previous phases, we mapped the learning progression across the IB continuum to identify points of alignment and misalignment between the *MYP Guide* and those of the PYP and DP by comparatively analyzing their treatment of language acquisition. The Progression Analysis, vetted by four levels of experts (i.e., IB classroom practitioners; school-, district-, and state-level administrators of IB programmes; IB development staff; and trained researchers with IB classroom experience), showed the *MYP Guide* to illustrate foundational principles of language acquisition and assessment that link well with the PYP and DP. Those principles include language learning that involves meaningful communication in authentic contexts and language acquisition that is valued for its ability to improve learners' intercultural competence. Programmatic differences include the terminology used, framework provided for language acquisition, and approaches to assessment.

² e.g., reviewers found that Objective A, Strand ii, did not vary meaningfully between Phases 1 and 2.

³ We elaborate upon the selection of international standard sets in the Method section.

Phase V: Discrepancy Analysis (Conclusion)

Examining data from the first four phases (representing contributions from MYP staff across various divisions and educators from IB schools in five countries) for convergence, complementarity, and contradiction raised four questions to guide deliberations for improving the *MYP Guide*:

1. How can the *MYP Guide* balance exemplification and full detail with simplification and clarity?
2. To what extent should language, assessment criteria, and other facets align across the IB continuum?
3. What assessment changes can the IB undertake without creating additional challenges for practitioners and schools?
4. What else can the *MYP Guide* borrow from the other international language frameworks?

Bolstered by findings from this mixed-methods design, it seems clear that the IB should attend to certain weaknesses in specificity and exemplification in a few areas of the *MYP Guide*. EPIC and CASLS recommend that IB begin with overarching conversations and consensus about the scope and purpose of change before making any decisions about what in the *MYP Guide* requires revision, how to approach that revision, and to what extent revision is advisable.

INTRODUCTION

The Educational Policy Improvement Center (EPIC), in partnership with the University of Oregon's Center for Applied Second Language Studies (CASLS), examined the alignment and coherence of the Middle Years Programme *Language Acquisition Guide's* (MYP *Guide*) description of language development for learners aged 11 to 16. The review involved five phases:

1. We *reviewed research literature* on language acquisition development and practices to examine how effectively the MYP *Guide* conceptualizes and describes a progression of learning an additional language.
2. We recruited and trained IB staff and school-based educators to conduct a *within-document analysis*, examining how effectively the MYP *Guide's* assessment criteria aligned with its other progressions of language learning.
3. We recruited and trained IB staff and school-based educators to conduct a *cross-document analysis*, examining the MYP *Guide* in comparison to well-known language acquisition frameworks from three international sources.
4. In our *progression analysis*, we identified points of alignment and misalignment between the MYP *Guide* and the language acquisition guides of the Primary Years Programme (PYP) and Diploma Programme (DP).
5. In a *discrepancy analysis*, we reexamined findings from the first four phases to see how they converged with, complemented, or contradicted one another. In this phase, we asked: What changes or refinements to the MYP *Guide* would promote good teaching practice and efficiency when evaluating student progress in language acquisition?

After the executive summary, this report presents (a) a detailed description of the methods that correspond to each phase, (b) itemized findings for each of the first four phases, and (c) a conclusion that presents the discrepancy analysis and four questions to guide IB's next steps.

In the ensuing sections of this report, we provide the methodological approach, findings, and the conclusion of this study. The conclusion includes findings-informed recommendations for potential revisions to the MYP *Guide*.

METHODOLOGICAL APPROACH

In this section, we discuss the methods used for each of the five phases of research in this project: Phase I: Literature Review, Phase II: Within-Document Analysis, Phase III: Cross-Document Analysis, Phase IV: Progression Analysis, and Phase V: Discrepancy Analysis. For each phase, we first report the research question(s) that propelled the phase and then outline the activities we used to address the question(s).

Phase I: Literature Review

We conducted a comprehensive literature review as we asked: *How effectively does the MYP language acquisition course conceptualize and describe a progression of learning?* The following subsidiary questions, provided by IB, guided our approach to collecting and analyzing literature that depicted learners' progressions in language acquisition:

- Do the MYP language acquisition course's four communicative processes (as expressed in objectives and assessment criteria) sufficiently reflect current thinking in educational research and practice with regard to language acquisition?
- Do the MYP's six phases of language acquisition reasonably reflect current thinking in educational research and practice with regard to language acquisition development?
- Is the course's global proficiency table precise, practical, and integrated with MYP phases and communicative processes?
- Are the course's language acquisition continuums (for listening and speaking, viewing and interpreting, reading comprehension, and writing) precise, practical, and integrated with the other MYP phases and processes?
- Does the subject-specific guidance for assessment task design align to other parts of the *MYP Guide*, which specify developmental progression?

Data Collection

To ensure that we used relevant resources to inform the most comprehensive review possible, we considered sources from the fields of language acquisition and education that met three criteria. Sources (a) were peer reviewed; (b) addressed listening and speaking, viewing and interpreting, reading comprehension, and writing; and (c) offered insights into how effectively IB assessment criteria map onto its progressions of language learning in the *MYP Guide* and/or how the developmental progression that it specifies can link with the language acquisition development progressions of the PYP or DP. Whenever possible, we included studies that directly addressed language learning processes of MYP-age learners (i.e., 11–16 years). However, given that IB devised the MYP language acquisition global proficiency table using the Council of Europe's (2002) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (CEFR) and the American Council on the Teaching of Foreign Languages' (ACTFL, 2011) *Standards for Foreign Language Learning in the 21st Century*—and neither of these documents specify MYP-age learners—it was necessary to expand our literature pool to include studies involving learners of various age groups. Furthermore, MYP ages are an under-represented bracket in second-language literature. To contend with this deficiency and provide a robust review, we considered all literature published since 1997 that met our criteria, but emphasized studies from the past 10 years.

Search Parameters and Data Analysis

As we investigated the appropriateness of the phases of language acquisition, we sought to evaluate the MYP *Guide*'s use of CEFR and ACTFL standards. To parse this research and its relation to the phases in the MYP *Guide*, we searched for articles that addressed phases or stages in language learning or second-language acquisition (SLA), which has been called L2 in some research literature.⁴ The search terms "phases in L2 acquisition," "phases in SLA," "phases in language learning," "stages in L2 acquisition," "stages in SLA," and "stages in language learning" did not yield many results, possibly because contemporary literature negates the assumption that language acquisition follows a linear progression (Westhoff, 2007). To find more relevant sources, we searched the same journals again, this time using the terms "validity ACTFL" and "validity CEFR." Those terms yielded a plethora of results that focused largely on the history of establishing the standards, the lack of empirical research to support the standards, and their degree of proliferation.

Initially, we evaluated the language processes in the MYP *Guide* for their potential to engage learners in meaningful communication (i.e., communication that produces and indicates language acquisition). Accordingly, we reviewed studies to identify the most important characteristics of meaningful communication. This search necessitated that we expand our temporal inclusion parameter to consider Canale and Swain's (1980) framework for communicative competence because it is foundational in language acquisition literature. This search also led us to consider literature pertaining to effective input, comprehension and interaction, and output. Then, we evaluated the theoretical underpinnings of the MYP *Guide*'s focus on visual interpretation in combination with reading, writing, listening, and speaking. Our evaluation included considerations related to Byram's (1997) model for intercultural competence, contextualized communication, and literature regarding multiliteracies and computer-mediated communication. Furthermore, given IB's focus on authentic, learner-centered assessment, we included Wiggins's original article (1989) on the need for authentic assessment and an opinion piece in which Wiggins (2011) reflects on how his principles have influenced pedagogical approaches to assessment over the past 20 years. We built on Wiggins's research with more contemporary views of authentic assessment to evaluate how context is embedded within assessed language processes.

Validation and Report of Findings

We followed a seven-step approach to validate findings from this phase. First, CASLS produced a topic-sorted annotated bibliography. Second, CASLS recruited three second-language acquisition research experts⁵ to review that bibliography and ensure that no critical studies had been omitted. Third, EPIC's researchers reviewed the annotated bibliography to ensure its fitness for purpose. Fourth, after we received approval from IB to continue, CASLS completed a full-length review of research findings as they linked to the MYP's four communicative processes, six phases of language acquisition, four skill areas, and assessment task design. Fifth, CASLS had the three experts read the review to ensure its technical accuracy. Sixth, EPIC's researchers examined the

⁴ In that instance, L1 would indicate an individual's mother tongue.

⁵ All three experts have PhDs, as well as active teaching and research agendas, in second-language acquisition. Collectively, the three reviewers offer additional expertise in multicultural education, computer-assisted language learning, interlanguage pragmatic development, L2 phonetics and phonology, and technology in foreign language education.

review to ensure its fitness for purpose. Seventh, we submitted to IB the full-length review so its staff could examine the findings. In this final report, we report highlights from the review, per IB's request. We also have included the annotated bibliography (Appendix A) and full-length review (Appendix B) as appendices.

Phase II: Within-Document Analysis

We recruited and trained IB staff and school-based educators to review the *MYP Guide* as our approach to addressing the following question: *How effectively do the MYP language acquisition course's assessment criteria map onto its other progressions of language learning in the Guide?*

Recruitment and Training of Reviewers

In consultation with IB, we recruited six reviewers: three members of the MYP staff and three MYP educators with expertise in teaching language acquisition. We sought reviews from IB staff and MYP educators to establish face validity and to capture the content aspect of construct validity. Face validity is achieved when a direct measure provides information that, on its face, responds to the question of interest (Nevo, 1985). The content aspect of construct validity requires evidence of content relevance, representativeness, and technical quality (Messick, 1995). To enhance the generalizability of findings, we purposively sampled one MYP educator from each of IB's three regions: (a) Africa, Europe, and the Middle East; (b) the Americas; and (c) Asia-Pacific. IB schools typically instruct in English, French, or Spanish, so sampling across IB regions generated contributions from a wider range of learned school languages (e.g., Chinese).

We developed a tool in Microsoft Excel to standardize the within-document analysis (see Appendix C). Using WebEx, we trained all six reviewers to use that tool. One MYP staff whom we recruited and trained for the within-document analysis could not complete the task due to scheduling demands. Therefore, this phase includes findings from five reviewers. Our five within-document analysis reviewers also completed a cross-document analysis (see Appendix D), for which we added another MYP staff for that phase of the project only. We describe that methodological process in a subsequent section of this report.

Review Tool and Analytical Procedures

Guided by findings from the full-length literature review, EPIC and IB together constructed a tool that drew on Porter's (2002) content matrix design. That design enabled our reviewers to work individually to indicate the document's levels of clarity, alignment, and appropriateness as necessary for eight distinct analytical tasks.

Clarity meant a reviewer knew what evidence of a component in the *MYP Guide* would look like in practice. Reviewers analyzed four components of the *MYP Guide* for clarity: assessment criteria, phase-specific language acquisition objectives, the global proficiency table, and the language acquisition continuums.

Alignment meant a reviewer interpreted a component in the *MYP Guide* to be the same or similar enough to a component of the relevant assessment criteria rubric. Reviewers analyzed three components for alignment: phase-specific language acquisition objectives, the global proficiency table, and the language acquisition continuums.

Appropriateness meant learners could meet the expectations that corresponded to the MYP *Guide*'s highest achievement level (i.e., 7–8). Reviewers analyzed phase-specific assessment task guidelines for appropriateness.

After each of the eight analytical tasks, we encouraged reviewers to amplify or specify their responses, or suggest problems, challenges, or recommendations. The review tool included a summative open-ended item that asked: *How effectively do the different progressions help you judge student placement and plan for instruction across the middle years?*⁶

Report of Findings

In the section on Findings, we first report interrater reliability statistics for each of the eight analytical tasks that composed the within-document analysis. Next, we report quantitative and qualitative findings by task. Quantitative findings include the proportions of reviewers who determined that a component was *Clear*, *Aligned*, or *Appropriate*, depending upon the given task. Qualitative findings include reviewers' specifications, suggested problems, challenges, or recommendations. We also report reviewers' responses to the open-ended item.

Phase III: Cross-Document Analysis

We recruited and trained IB staff and school-based educators to conduct independent cross-document analyses of the MYP *Guide* along with language frameworks from three international sources as our approach to addressing the following question: *How do the progressions of learning compare with known standards of language acquisition?*

Recruitment and Training of Reviewers

The same three MYP educators and two of the IB staff who conducted the within-document analysis also conducted the cross-document analysis. IB selected a third individual from its staff to complete the review team. Again, we developed a tool in Microsoft Excel to standardize the cross-document review process. Using WebEx, we trained all six reviewers to use that tool to examine the MYP *Guide* and one of three international language frameworks individually, in comparison to research on language acquisition. We intentionally sequenced the cross-document analysis after the within-document analysis to capitalize on most reviewers' in-depth familiarity with the MYP *Guide*, having already reviewed it during the within-document analysis.

Data Sources and Analytical Procedures

All six reviewers rated the MYP *Guide*, and two reviewers were each assigned to review one of the following international sources: The Council of Europe's *Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (CEFR); the American Council on the Teaching of Foreign Languages' *Standards for Foreign Language Learning in the 21st Century* (ACTFL); and the Hong Kong Education Bureau's *Key Learning Areas: English and Chinese Language Education* (HKEB).⁷ Using a four-level scale (i.e., *not at all*, *too little*, *enough*, *too much*) to assess the amount of guidance

⁶ We also asked reviewers: *What changes would you like to see that would improve the usefulness of the MYP Language Acquisition Guide to educators?* Per agreement with IB, we do not report findings from that item here. Instead, we included the item's raw data in our previously submitted interim report.

⁷ We selected these three standard sets in consultation with IB to generate the most geographically representative sample possible.

provided in the *MYP Guide*, each of the six reviewers rated the document on 37 components of language learning that we derived from our full-length literature review on language acquisition instruction. The 37 components fit within eight domains:

- Interpersonal communication
- Defining phases of language acquisition
- Communicative competence and communicative processes
- Effective input, comprehension and interaction, and output
- Intercultural competence
- Contextualized communication
- Multimodality and multiliteracies
- Assessment

Due to the differences in the numbers of reviewers per document, EPIC defined consensus as (a) when 5 of 6 or all six reviewers of the *MYP Guide* indicated the same rating and (b) when both of the CEFR, ACTFL, or HKEB reviewers indicated the same rating for their respective standard set. Thus, EPIC designed its analyses to neither ignore outlier ratings nor accord too much weight to any single discrepant view, in line with Herrera-Viedma, Cabrerizo, Kacprzyk, and Pedrycz's (2014) unanimity-minus-one approach to decision making.

Last, each reviewer completed two open-ended exit interview items. We asked all six reviewers: *Are there unique strengths within the MYP Guide's progressions of learning compared to the other framework of language learning? If so, please describe these unique characteristics below.* Then we asked each pair that shared a given international source: *Are there unique strengths within the [NAME OF INTERNATIONAL SOURCE] that, if incorporated into the MYP Guide, would enhance its usefulness for educators? If so, please describe these strengths and provide any suggestions on how they could be used to improve the MYP Guide.*

Report of Findings

In the section on Findings, we first report interrater reliability statistics for the reviews of each of the four documents, disaggregating reliability for the *MYP Guide* by reviewer's role (i.e., IB staff or MYP educator). Next, we report quantitative findings, which summarize overall and component-specific observations about the extent to which the reviewers rated the documents as providing too much, enough, too little, or no information about a given component. We conclude the Findings section for this phase by examining themes that emerged from the reviewers' responses to the open-ended items.

Phase IV: Progression Analysis

Drawing on findings from the previous phases of this project, we mapped the learning progression across the IB continuum to identify points of alignment and misalignment as we addressed the following question: *Does the developmental progression as specified in the MYP Guide link sufficiently with Primary Years Programme (PYP) and Diploma Programme (DP) language acquisition development?*

To address this question, we examined points of possible alignment and misalignment in the developmental progressions of language acquisition articulated in the PYP, MYP, and DP.

We addressed this topic by comparatively analyzing the treatment of language acquisition in MYP and DP subject-area guides and in two PYP documents: *Making the PYP happen: A curriculum framework for international primary curriculum* and *Language scope and sequence*.

Validation and Report of Findings

To validate our findings, we employed several levels of review.⁸ First, CASLS conducted the progression analysis and vetted its findings internally. Second, CASLS recruited three IB experts⁹ to examine the methods of analysis and findings from the progression analysis. Third, EPIC's researchers examined the methods of analysis and findings.

Report of Findings

In the Findings section, we present a synthesized table to summarize points of developmental progression with points of alignment and misalignment that we identified across phases, skills, and assessment tasks. We supplement that table with narrative to clarify findings.

Phase V: Discrepancy Analysis

In this final phase, we asked: *Are changes or refinements needed in the structure of the MYP Guide to promote good teaching practice and efficiency in placement and evaluation of student progress in language acquisition?* To address this question, we reflected on the distinct analyses from each of the previous phases to illuminate possible areas of convergence, complementarity, and contradiction (Greene, Caracelli, & Graham, 1989).

Data Collection

Our process included five activities. First, we sampled two MYP language acquisition teachers, one each from Argentina and Canada, for audio-recorded interviews. Our interviewees have between them experience instructing in IB's three main languages (i.e., English, French, and Spanish) and have worked in schools that offer the PYP, MYP, and DP. Second, we sent each of the interviewees the findings from Phase IV: Progression Analysis via email, allowing them a week's time to review the document. Third, we developed a semistructured interview protocol based on the first three phases of this research project (see Appendix E for protocol). Fourth, we interviewed the MYP language acquisition teachers individually about the extent to which the progression analysis conformed to their experiences working in IB continuum schools. Fifth, we analyzed findings from those interviews to develop a semistructured interview protocol for a final, audio-recorded interview with the Head of MYP Development at IB's request (see Appendix F for protocol). We also conducted a member check after an initial reporting of findings (Merriam, 1998).

Data Analysis and Report of Findings

We arrayed findings from the literature review, within-document analysis, cross-document analysis, progression analysis, and interviews during this phase. Examining findings comparatively across

⁸ We incorporated two additional levels of review for this analysis into Phase V.

⁹ All three experts have supervisory roles in public K–12 school jurisdictions that offer the full continuum of IB schools. Two of three experts have doctoral degrees. The expert without a doctoral degree serves as a member of the IB Educator Network and heads a state-level IB association. Across the experts, they have been responsible for overseeing authorization and/or implementation of all four IB programmes.

phases, we found some findings to converge with one another, some to complement one another, and a third group that contradicted one another. We organized findings in these three groups to serve as this study's Conclusion. For areas of convergence, we report illustrative results. In areas of complementarity, we show distinctions that overlap, but note areas where limitations of findings from one approach are mitigated by strengths of another. In areas of contradiction, we provide initial hypotheses regarding the causes or sources of the contradictions.

FINDINGS

We report findings in phase order from Phase I: Literature Review through Phase IV: Progression Analysis. Findings from Phase V: Discrepancy Analysis have been reserved for the Conclusion section. Due to the vast amount of data we collected and analyzed for this project, we begin each phase's subsection with one or more callouts to highlight key findings.

Phase I: Literature Review

Our literature review revealed that the *MYP Guide* has solid theoretical underpinnings, but presents a few practical challenges. As strengths, the *MYP Guide* directs educators to provide learning opportunities for students in which they are engaged in meaningful communication, fostering the growth of their respective language proficiencies. Relatedly, the *MYP Guide* shows communication accurately, as an activity that is complex, multimodal, and integrative. As currently constituted, however, the *MYP Guide* yields some confusion that might thwart application in classrooms. The *MYP Guide's* language processes and continuums do not align directly with CEFR and ACTFL, two standard sets that are internationally ubiquitous—despite a limited research base for their validity. On one hand, misalignment with these standards might be particularly problematic for a multinational network of schools such as IB, above and beyond the challenges it might present for an individual

Strengths

- Focusing on communicative competence rather than low-level skill
- Effective input/output
- Intercultural competence
- Multimodality

Challenges

- Clarifying language continuums and processes
- Assessing interpretation and production together
- Conceptualizing classrooms based on proficiency

school in a single national context. On the other hand, neither the CEFR nor the ACTFL standards have been subjected to particularly rigorous validation efforts. Therefore, the current project might position the *MYP Guide* to provide as valid, if not a more valid, approach than the internationally renowned standards for language acquisition. Furthermore, the *MYP Guide's* emphasis on multimodality and integrative assessment tasks complicates the already arduous task that educators face: providing comprehensible feedback. Meanwhile, the *MYP Guide* offers little explicit direction for educators to integrate meaningful communicative tasks daily into their classrooms.

Although it should be noted that IB's teacher support materials unpack further some of the expectations in the *Guide* because it is not meant to offer explicit direction given the myriad contexts that IB schools inhabit. Our review highlighted six areas that warrant further discussion and three recommended changes that could maximize the *MYP Guide's* usefulness for educators.

Highlights

Second-language acquisition is interdisciplinary, complex, and difficult to describe (Gass & Selinker, 2008). Our literature review serves to bring some cohesion to the considerably broad research base on language acquisition. We highlight six overarching themes:

1. **Phases of language acquisition:** Following modern language progressions such as those found in CEFR and ACTFL, the *MYP Guide* casts proficiency across six increasingly

complex phases. This approach is generally accepted, but not yet supported empirically (Hulstijn, 2007; Hulstijn, Schoonen, Jong, Steinel, & Florin, 2011; Liskin-Gasparro, 2003; Little, 2007; North, 2007). Furthermore, though a progression approach is the norm, scholars have not reached consensus on which progression is best (Gass & Selinker, 2008; Hulstijn et al., 2011). Moreover, the proliferation of CEFR and ACTFL has crowded out the possibility of other options (Alderson, 2007; Liskin-Gasparro, 2003).

2. **Communicative competence:** The *MYP Guide* does not overemphasize lower-level skills such as grammatical competency¹⁰ (Devitt, 1997), instead focusing on higher-level skills that enable language learners to process authentic, meaningful input and engage in authentic, meaningful output tasks. The literature clearly regards this approach as a strength.
3. **Effective input, comprehension and interaction, and output:** Prompting learners to analyze a variety of text types empowers them to communicate within and beyond classroom contexts (Hyland, 2007), a core input that the *MYP Guide* emphasizes through its assessment criteria. Moreover, the *MYP Guide* seeks to compel learners into interactions in which they must negotiate meaning as they acquire language (see Yuksel & Inan, 2014). The *MYP Guide* achieves this aim by steering educators toward tasks for learners that simultaneously build and prove their language knowledge (Mantero, 2002).
4. **Intercultural competence:** The interplay of the *MYP Guide's* continuums and assessment criteria / communicative processes promotes intercultural competence: the attitudes, skills, and knowledge to revisit personal values and beliefs when confronted with disparate beliefs without valuing one over the other (Byram, Nichols, & Stevens, 2001). Skills to interpret and relate, discover and interact, and possess critical cultural awareness align with the IB mission and core notion of international-mindedness.
5. **Contextualized communication:** The *MYP Guide* demonstrates the theoretical strength of blending its language continuums into processes, a recognition that language acquisition is both socially mediated and cognitive (Atkinson, 2002).
6. **Multimodality and multiliteracies:** The *MYP Guide* justifiably supports exploration of digital forms of communication, an increasingly evident form of input and platform for output that learners experience daily both within educative contexts and in extramural contexts (Lotherington & Jenson, 2011).

Recommendations Based on Literature Review

Though our review shows the *MYP Guide* to be theoretically sound, we identified three potential solutions to some of its complications around implementation.

1. **Clarify the selection of language continuums, language processes, and how they interact.** Unlike the CEFR and ACTFL standards, the *MYP Guide* focuses more on integrative communication, combining listening and speaking as one continuum and requiring interpretation and production for assessment criteria in most instances. Furthermore, IB separates visual and interpreting skills into their own continuums; CEFR and ACTFL embed those modes. Given these departures from the ubiquitous

¹⁰ In this report, we use the term *competence* to describe an overarching area of ability. We use the term *competency* to describe a specific skill that, along with other competencies comprise one's overall competence, in a given domain. We recognize that the academic literature related to second-language acquisition is not uniform in its use of terms such as competence and competency. We distinguish these terms as described previously to minimize confusion for readers of this report.

standard sets, IB should consider altering the continuums and processes to more closely reflect the international standards or create support materials that show educators why IB selected these continuums and processes and how they align with or improve on those international standards.

2. **Require learners to engage in both input and output on a single assessment to avoid complicating feedback loops.** The *MYP Guide* expects learners to interpret and produce language in 3 of 4 processes. Interpretive and productive forms of communication rarely exist without influencing one another. This theoretically sound approach can provide necessary context to shaping communication, but it also creates situations whereby rubric descriptors might be too wide-ranging to inform learners of their progress. Learners need to understand their capabilities and deficiencies to improve; noticing one's capabilities and deficiencies becomes complicated by the distinct macrolevel skills being evaluated within each assessment. To provide clear feedback to learners while maintaining an integrative approach, IB could design an assessment cycle that resembles ACTFL's Integrated Performance Assessment (IPA).¹¹ In executing an IPA, learners receive a series of tasks within each mode of communication and aligned to a single theme. Learners are evaluated for their performance in each mode but receive feedback along the way, a process that allows for integration and a clear loop that informs learners of their capabilities and deficiencies.

3. **Provide guidance to help educators conceptualize a proficiency-based classroom.** The *MYP Guide* appropriately advocates that language classrooms center on developing language proficiency rather than acquiring knowledge about language. But despite abundant evidence from the *MYP Guide's* integrative and contextualized communication tasks, there is little information for how educators could use similar tasks formatively rather than as proof that learning occurred. Including examples of daily classroom tasks mapped in a progression toward a summative assessment of a given communicative process would improve the *MYP Guide's* usefulness.

Reviewer-Identified Concerns

- Imprecision / inconsistency in use of terms
- Desire for illustrative examples
- Lapses in logic
- Need to differentiate between strands
- Lack of explicit definitions

Phase II: Within-Document Analysis

In Table 1, we report interrater reliability statistics for each of the eight analytical tasks. In 2-of-8 tasks, the roles of the reviewers seem to have had an influence, albeit minor, on reliability calculations. For the assessment criteria reviews of clarity, the two IB staff reviewers agreed in 72% of decisions; the three educators agreed at a rate of 62% for the corresponding decisions. For the phase-specific language acquisition objective reviews of clarity, the three educators agreed in 87% of decisions; the IB staff agreed at a rate of 62% for the corresponding decisions. Neither instance had a meaningful impact on the reliability above chance agreement calculations that we report in Table 1. For the other six analytical tasks, there was no evidence of role-related effects on reliability. After discussing reliability, we present quantitative and qualitative findings from each analytical task per the task order in Table 1.

¹¹ For more information, see http://carla.umn.edu/assessment/vac/CreateUnit/p_2.html

Table 1. Reviewer’s Interrater Reliability Statistics by Task

Analytical task	Decisions	Average pairwise agreement (%)	Reliability above chance agreement (α)
Assessment criteria phase-specific descriptions (Clarity)	600	54	.04
Phase-specific language acquisition (Clarity)	390	75	.05
Phase-specific language acquisition (Alignment with assessment criteria)	120	65	-.09
Global proficiency table (Clarity)	30	87	-.04
Global proficiency table (Alignment with assessment criteria)	120	75	-.07
Language acquisition continuums (Clarity)	240	85	-.05
Language acquisition continuums (Alignment with assessment criteria)	120	100	Undefined
Phase-specific assessment task guidelines (Appropriateness)	120	85	.32

Note. Average pairwise agreement is the mean percentage agreement of each possible pair among the five reviewers. Reliability above chance agreement (i.e., Krippendorff’s alpha) shows the five reviewers’ level of agreement above and beyond the possibility of agreeing at random; undefined values occur for reliability above chance calculations if pairwise agreement is 100%.

Assessment Criteria Phase-Specific Descriptions (Clarity)

In Figures 1–4, we report the number of reviewers who rated assessment criteria phase-specific descriptions as Clear for a given achievement level and phase. Figure 1 depicts trends for Criterion A (comprehending spoken and visual text); Figure 2 depicts trends for Criterion B (comprehending written and visual text); Figure 3 depicts trends for Criterion C (communicating in response); and Figure 4 depicts trends for Criterion D (using language in spoken or written form).¹²

All reviewers found achievement level 0 to be Clear across almost all assessment criteria phase-specific descriptions. For Criterion A, the number of reviewers rating Clear increased from Phases 1–6 in general for achievement levels 1–8. Reviewers’ numbers of clarity ratings for Criterion A were identical for achievement levels 1–2 and 5–6, so their trend lines overlap in Figure 1.

¹² Hereafter, criteria referenced in the following shorthand: Criterion A (comprehension spoken/visual); Criterion B (comprehension written/visual); Criterion C (communicate in response); Criterion D (use language in spoken or written form).

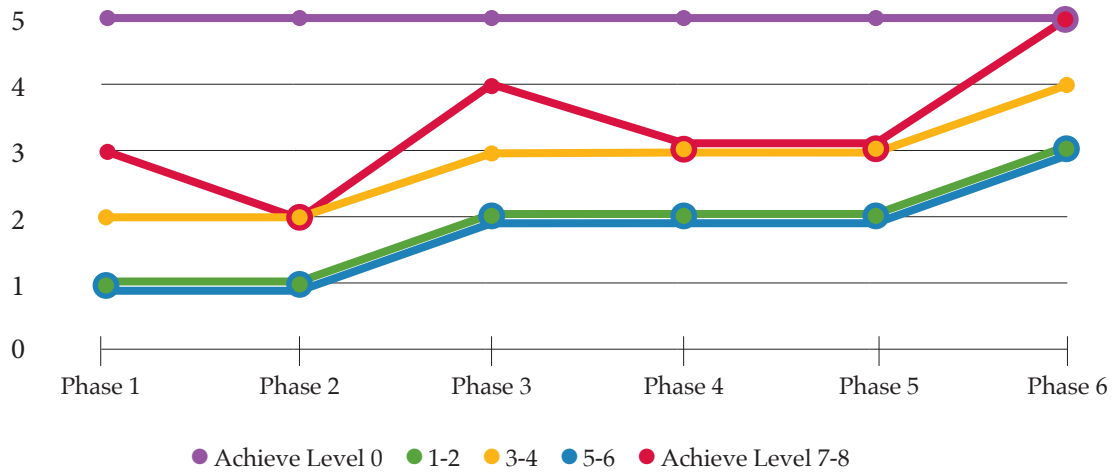


Figure 1. Number of Reviewers (N = 5) rating Comprehension Spoken/Visual (Criterion A) as Clear by phase and achievement level.

Regarding Criterion B, trend lines for reviewers' clarity ratings were relatively flat across phases (see Figure 2). Two of five reviewers provided a rating of Clear for all phases for achievement levels 1–2 and 5–6, so their trend lines overlap in Figure 2. Three of five reviewers provided a rating of Clear for all phases for achievement level 3–4. Four of five reviewers rated Clear for Phases 2–6 for achievement level 7–8.

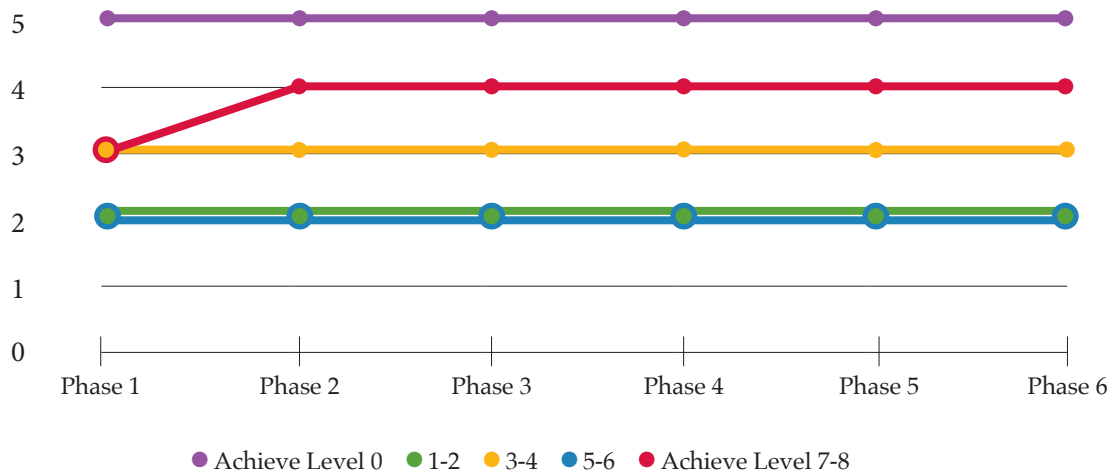


Figure 2. Number of Reviewers (N = 5) rating Comprehension Written/Visual (Criterion B) as Clear by phase and achievement level.

Communicate in Response (Criterion C) was the only criterion in which a reviewer rated achievement level 0 as *Not Clear* (see Figure 3). Reviewers' ratings for Criterion C across phases produced overlapping trend lines for achievement levels 1–2 and 3–4 and for achievement levels 5–6 and 7–8.

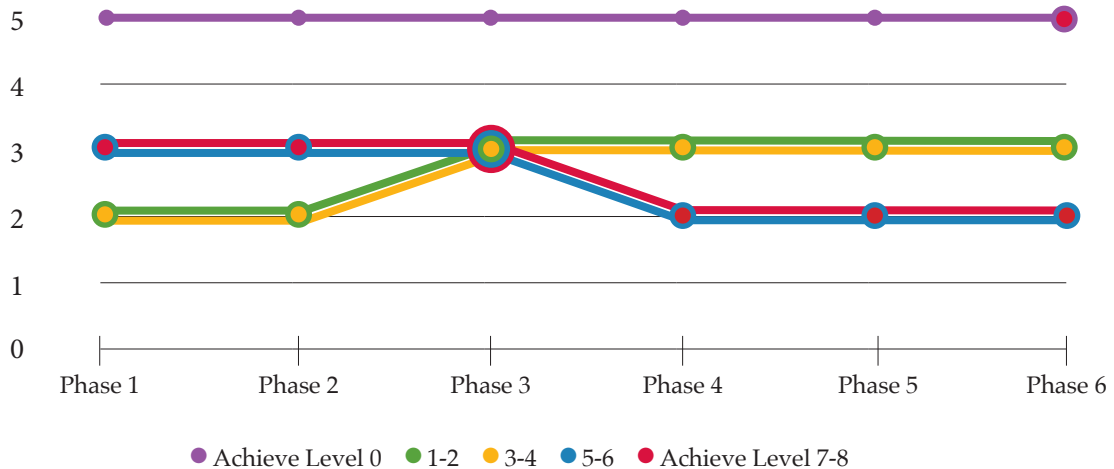


Figure 3. Number of reviewers ($N = 5$) rating Communicate in Response (Criterion C) as Clear by phase and achievement level.

Regarding Criterion D, 2 of 5 reviewers rated Clear for all phases for achievement levels 1–2 and 3–4. For achievement levels 5–6 and 7–8, reviewers' clarity ratings showed overlapping trend lines for Phases 1–5.

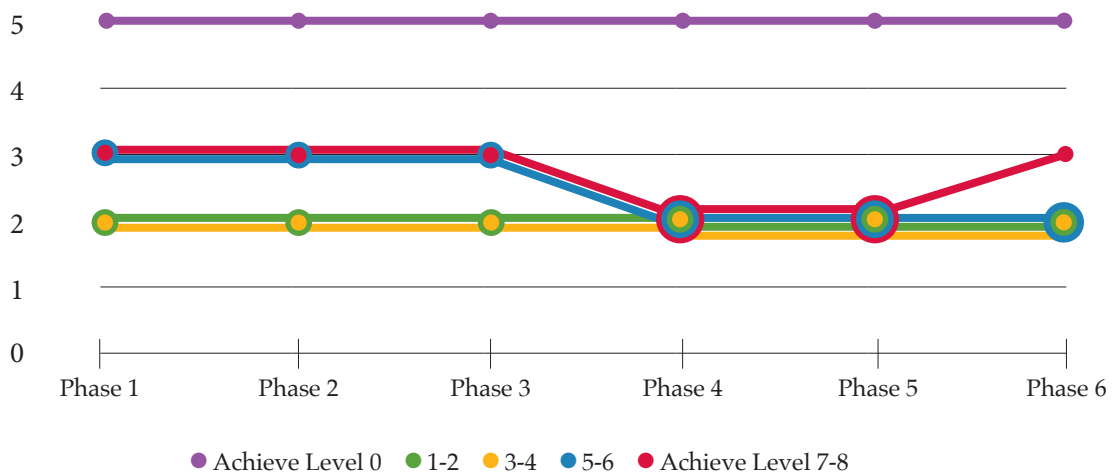


Figure 4. Number of reviewers ($N = 5$) rating Use Language in Spoken or Written Form (Criterion D) as Clear by phase and achievement level.

The qualitative data we generated from reviewers' amplifications and our exit interview item provide some explanation for the patterns we observed and for the data that did not follow distinct patterns. However, it is important to note that our two IB staff found clarity to be lacking in far more decisions around clarity for the assessment criteria phase-specific descriptions (81% and 75%, respectively) than did our three MYP educators in the same analytical task (44%, 40%, and 0%, respectively). The problems, challenges, and recommendations identified by reviewers formed three categories. First, imprecise use and/or inconsistent application of terminology reduced clarity in some aspects of the MYP *Guide*. Second, educators could benefit from examples that illustrate expectations in the MYP *Guide*. Last, questions arose about the appropriateness and/or logic that undergird components of the MYP *Guide*.

Imprecise use and/or inconsistent application of terminology. Four of the five reviewers found at least one term to be used imprecisely or applied inconsistently. As one educator noted, the MYP *Guide* clarifies its use of "conventions" as "format, style and author's purpose" in Criterion B, but does not do so for Criterion A. Moreover, another educator discounted the need for a general description in receptive language skills' Criteria A and B about "making a response to the text based on personal experiences and opinions." According to that educator, learners confronted with these tasks "don't have to draw conclusions" but "only have to understand the text." By contrast, Criteria C and D carry expectations of creating responses. Therefore, the educator suggested that IB exchange "making" for another verb that does not "imply the creation of a response" within Criteria A and B. Another educator focused on Criteria A and B, specifically achievement level 1–2 in Phases 4–6. The educator advocated replacing "has difficulty" with terms such as "limited" and/or "few" to align more closely with other expectations in those assessment criteria phase-specific descriptions. Also in Criteria A and B, one IB staff questioned the denotation of the Phase 4 phrase "construct meaning" for being difficult to define and not lending itself to measurement.

In Phases 4–6 for Criterion C, an educator found difficulty distinguishing "considerably" in achievement level 5–6 from "confidently" in level 7–8. The educator preferred if both achievement levels would expect "appropriate" and "accurate" responses, terms used in Criterion D, Phase 4. Still, the same educator identified an overlap between appropriate and accurate, seeking clearer, more consistent use of those terms. The same educator wondered if "inaccurate" would be more useful than "inappropriate" for achievement levels 1–4 because "appropriateness is very subjective, whereas accuracy is more measurable." The educator also wondered whether responses "could be inappropriate without also being inaccurate."

One educator called out an inconsistency across the learning continuums. For all phases, Criterion D, Strand i, mentions speaking, but neglects writing in a criterion meant to assess learners' uses of language in both spoken and written forms. In one review from IB staff, a related observation in Criterion C, Strand i, wondered if the title at Phase I should refer to spoken, written, and/or visual text, because the achievement levels mention all those skills. The same IB staff review called for a higher degree of overall alignment between Criteria A and B, specifically how Strand i expects Phase 2 learners to draw conclusions in Criterion B, but not Criterion A, and how Strand ii expects Phase 3 learners to "understand basic conventions" for Criterion B, but only "understand conventions" for Criterion A. Relatedly, the IB staff review noted that the phrase "in social and academic situations" appears atop achievement levels in Criterion A, Strand ii, Phase 6, but not in the strand itself, the lower achievement levels, or anywhere in Criterion B. However, it does appear in achievement level 7–8 of Criterion C.

Illustrative examples. Each reviewer identified at least one instance in which educators using the *MYP Guide* would benefit from illustrative examples. One educator noted that Strand ii across Phases 1–5 characterizes “basic conventions” too generally because educators “have their own interpretation” of basic conventions. The educator appreciated the clarity that Strand ii in Phase 6 provides on this issue. Another educator broadly called for examples across all phases, achievement levels, and assessment criteria phase-specific descriptions. More specifically, reviews from an educator and IB staff cited difficulty in envisioning what should constitute terms such as minimal, limited, adequate, some, considerable, usually, to some degree, most, and thorough. The IB staff review cited an inability to exemplify these terms as the main reason for nearly all her Not Clear responses to this analytical task. Another IB staff review concurred, describing some strands as contributing “to a general lack of clarity in all the achievement levels for all assessment criteria [phase-specific descriptions].” IB staff expected that the qualitative terms (e.g., minimal, limited) would yield various interpretations among educators, suggesting instead a glossary to standardize meanings for those terms. Focusing on Criteria A and B as the achievement levels move from 1–2 to 3–4, an educator identified the same concern, but suggested instead using “limited” across achievement levels but adding level-relevant nouns after limited (e.g., limited basic facts) to spare educators from having to interpret potentially minute distinctions between adjectives.

Appropriateness and logic. Three reviewers—two of whom were educators—offered several examples in which they questioned the appropriateness of the *MYP Guide*. One educator wondered if Criterion B, Phase 1, about author’s purpose asked “too much” of learners, and might be better suited for later phases. The same educator questioned Criterion C, Phase 1, based on an expectation that emergent learners would struggle to vary their communications to indicate senses of different audiences without having first acquired basic vocabulary. IB staff suspected a Phase 2 learner should not be expected to have “excellent understanding of messages, main ideas and supporting details.” IB staff also questioned (a) overlaps in phases in Criterion A, Strands ii and iii, and Criterion B, Strand iii; (b) whether assessing students’ awareness of audience as “some, considerable, etc.” was appropriate for Criterion C; (c) how level of complexity might vary from Phases 3–6 for Criterion C, Strand i; (d) if confidence as stated in Criterion C, Strand ii, should determine assessment of ability; or (e) why Criterion D, Strand ii, assesses the amount of information learners can organize rather than organizational abilities (i.e., ordering ideas/information properly or systematically).

An educator and IB staff raised questions about some of the logic that undergirds the *MYP Guide*. The educator interrogated why Criterion B for Phases 3 and 4 allows learners to achieve level 1–2 if they cannot draw “any conclusions”: the stated objective requires learners to draw conclusions. The educator offered a change in which achievement level 1–2 would expect “limited or few conclusions.” Additionally, IB staff cited “lengthy discussions among teachers” in which they indicated confusion about whether Criterion C, Strand ii, truly assessed nonverbal communication among Phase 1 learners.

Phase-Specific Language Acquisition Objectives (Clarity)

In Figures 5–8, we report the number of reviewers who rated phase-specific language acquisition objectives as Clear for a given criterion, strand, and phase. Figure 5 depicts trends for Criterion A (comprehension spoken/visual), Figure 6 depicts trends for Criterion B (comprehension written/visual), Figure 7 depicts trends for Criterion C (communicate in response), and Figure 8 depicts

trends for Criterion D (use language in spoken or written form). All reviewers found Criterion D, Strand iii, to be Clear across all six phases. For all but three strands (A-ii, A-iii, and C-ii), reviewers agreed on clarity for Phase 3. There were no observable patterns for other strands or phases. As for the previous analytical task, we supplement descriptive statistics with qualitative findings and note that IB staff (24% and 20% of decisions, respectively) rated phase-specific objectives as Not Clear more often than did educators (8%, 5%, and 3%, respectively).

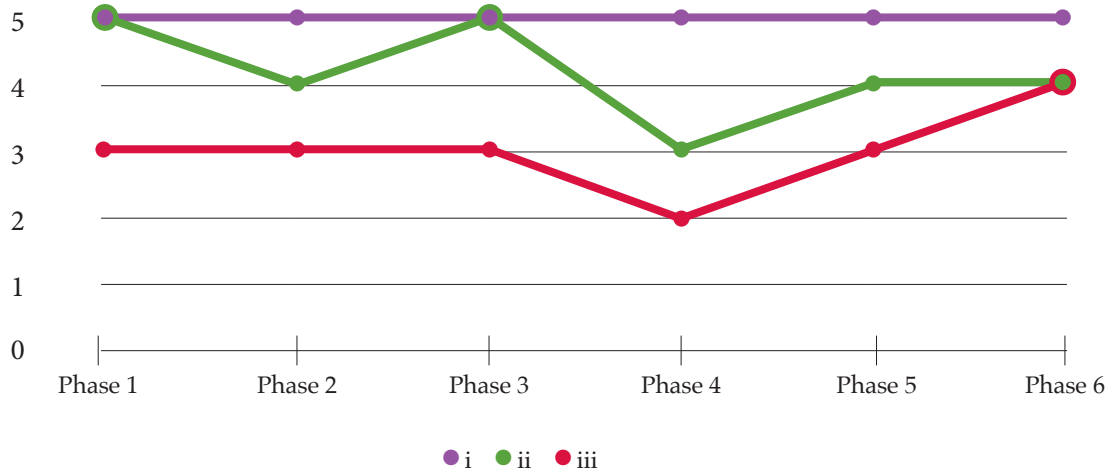


Figure 5. Number of reviewers (N = 5) rating phase-specific language acquisition objective as Clear by phase and strand for Criterion A.

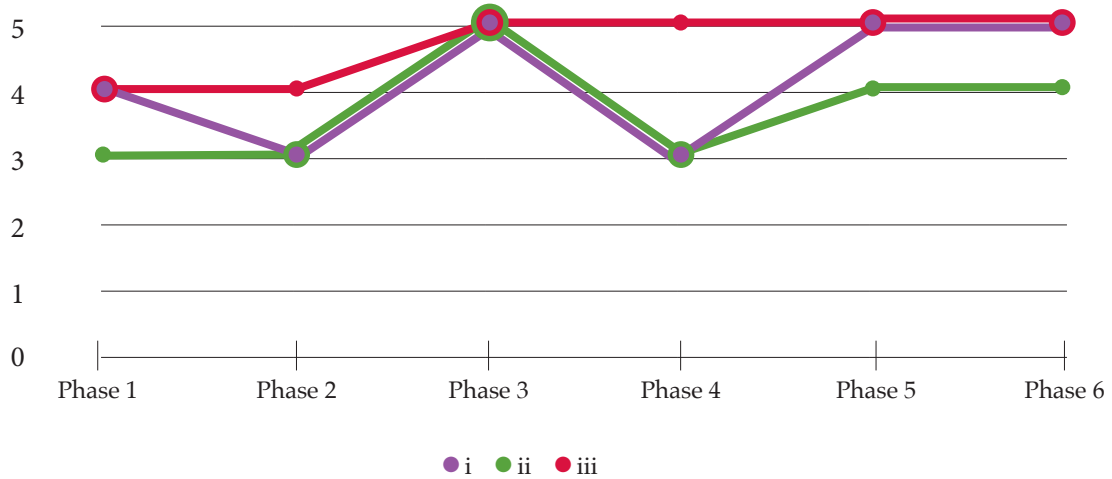


Figure 6. Number of reviewers (N = 5) rating phase-specific language acquisition objectives as Clear by phase and strand for Criterion B.

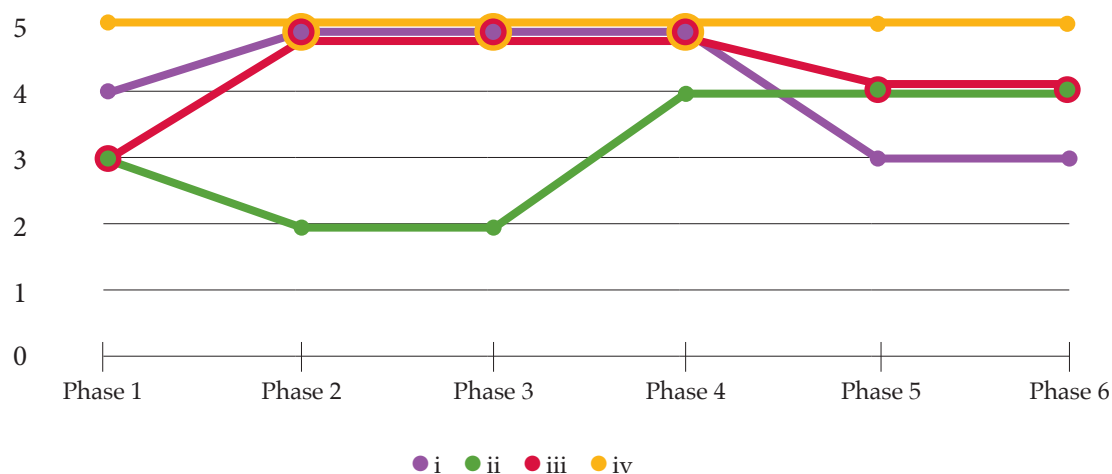


Figure 7. Number of reviewers (N = 5) rating phase-specific language acquisition objective as Clear by phase and strand for Criterion C.

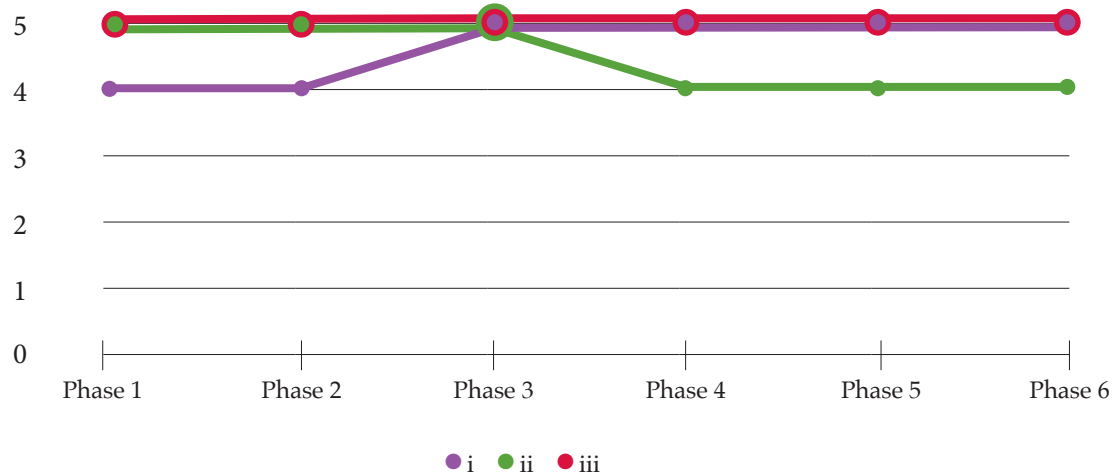


Figure 8. Number of reviewers (N = 5) rating phase-specific language acquisition objective as Clear by phase and strand for Criterion D.

Stemming from this analytical task, our reviewers identified three types of situations that would benefit from revision: (a) differentiating strands from one another more sharply, (b) defining terms within a given strand more explicitly, and (c) applying more appropriate and/or consistent logic to the use of terms.

Differentiation. Two IB staff and one educator identified points at which the phase-specific language acquisition objectives could be clarified through greater differentiation between strands. IB staff noted that Objective A, Strand ii, did not vary meaningfully between Phases 1 and 2. The educator preferred Objective C, contrasting it with the lack of variation accompanying the personal responses to text in Objectives A and B, as well as intonation in Objective D. Furthermore, the educator highlighted Objective D, Strand ii, noting issues for Phases 3–6. The educator

suggested clearer separation from Phase 2 such that Phase 3 would expect learners to “organize information and ideas *clearly*,” expect Phase 4 learners to “organize information and ideas *clearly and coherently*,” and expect Phase 5 learners to “organize information and ideas *into a structured text*” before Phase 6 learners would organize information and ideas clearly and express/compare them “*with personal ideas*.” Relatedly, the educator wondered why Phase 4 features a structured text if Phases 5 and 6 do not.

IB staff identified many phases that could feature greater differentiation. For Strand iii of Objectives A and B, IB staff cast Phase 1 as identical to Phase 2, the same degree of overlap that can be found between Phases 3 and 4 and Phases 5 and 6. IB staff made the same observation for all three pairs of phases of Objective C, Strand iv, and for Phases 1 and 2 and Phases 3 and 4 of Objective D, Strand i. IB staff noted that Phases 5 and 6 were still “very similar” for that strand. In addition, IB staff found Objective D, Strand ii, to be the same for Phases 2 and 3 and Phases 5 and 6. Last, IB staff characterized the wording for Objective D, Strand iii, as “the same at all six phases.”

Explicit definitions. Both IB staff and two educators recommended instances for the *MYP Guide* to offer more explicitly defined terms. IB staff both identified Criterion A as defining “conventions” in Phase 6, but the term remained undefined in earlier phases. One IB staff review argued for a definition in Phases 1 and 2, expressing concern that learners would be unprepared to meet Phase 3 expectations otherwise. An educator concurred, noting “quite a lot of teachers still didn’t get the term ‘conventions,’” so defining the term would be beneficial, especially given differences that the educator perceived between the term as the *MYP Guide* applies it in Objective A, Strand ii, versus its application in Objective D, Strand i.

Reviewers called out several other terms that could benefit from more explicit definition such as “construct meaning,” pondering if that phrase might or might not differ from “analyzing or drawing conclusions” in Objective A, Strand i, Phase 4. An educator remained unclear about how learners providing evidence as an interpretation of conventions in Objectives A and B, Strands ii, Phase 4, might contrast from an analysis of conventions in Objective B, Strand i, Phase 4. Furthermore, the educator wondered about the meaning of “aspect” of a convention in Objective B, Strand ii, Phase I, and sought clarity about what a “basic aspect” of format, style, or author’s purpose might look like. An IB staff review echoed this gap. The educator suggested substituting topics for aspects. Last, the educator asked if the phrase “basic structured exchange” implied rehearsed or unrehearsed scenarios in Objective C, Strand ii, Phases 2 and 3.

One IB staff review interrogated the term “engage” as used in Objective A, Strand iii, for Phases 1 and 2, imagining that it suggested “some deep understanding which students of this phase may not be able to demonstrate.” The same IB staff review raised this issue again in a critique of Objective B, Strand iii. Moreover, the IB staff review wanted to know if “interaction” as used in Objective C, Strand ii, featured oral, written, or other skills.

Appropriateness and logical consistency. Only IB staff raised issues of appropriateness and logical consistency for this analytical task. One IB staff review pointed to Objective D, Strand i, as being “heavy on language requirements especially for the lower phases.” Another IB staff review cited inconsistent uses of terms across Strand i, Phase 2, such as “identify” in Objective A and “show meaning of” in Objective B, when the same construction would lead to less confusion between objectives. The same IB staff review noted the command terms for Objectives A and B, Strand ii, alternate from “interpret” to “analyze” and back to “interpret” across Phases 4–6.

Also, in Objective C, Strand i, Phase 1, the IB staff review asked why “spoken, written, and/or visual” had been omitted. That IB staff review found cumbersome wording in Objective C, Strand ii, Phases 1–5, and Objective C, Strand i, Phase 1. Last, the IB staff review found inconsistent uses of the terms “social and academic situations” and/or contexts, at times accompanied by “familiar and unfamiliar” and at times not (see Objectives A, B, and C, Strands i and iii, for Phases 4–6).

Phase-Specific Language Acquisition Objectives (Alignment)

In Table 2, we report the number of reviewers who rated phase-specific language acquisition objectives as Aligned to each assessment criterion and phase. Across phases, 3 of 5 reviewers found alignment for Criterion C (communicate in response) and 4 of 5 reviewers found alignment for Criterion D (use language in spoken or written form). For Criterion B (comprehension written/visual), 4 of 5 reviewers found alignment for Phases 1 and 2; we had consensus for Phases 3–6. Criterion A (comprehension spoken/visual) showed more variation. In the next section, we report (in phase order) alignment problems or challenges that reviewers found in the objectives and/or their recommendations for those problems or challenges.

Table 2. Number of Reviewers (N = 5) Rating Phase-Specific Objectives as Aligned by Phase

Assessment criterion	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6
A. Comprehension spoken/visual	4	4	4	5	5	4
B. Comprehension written/visual	4	4	5	5	5	5
C. Communicate in response	3	3	3	3	3	3
D. Use language in spoken or written form	4	4	4	4	4	4

Problems, challenges, and recommendations. Four of the five reviewers reported at least one problem, challenge, or recommendation for this analytical task. For Phase 1,¹³ an educator noted the challenge that Objective C, Strand i, expects learners to “respond appropriately to simple short phrases,” assessment criteria phase-specific descriptions of other achievement levels in Phase 1 expect responses to be “short phrases, but also basic information in spoken, written and visual texts.” Another educator recommended a change for Criterion B: Shift the use of the term excellent for a more “appropriate” phrase (e.g., “has basic awareness”). For Phases 1 and 2, an IB staff review noted that Criterion B, Strands i and ii, do not “match the learners’ levels” and that the MYP *Guide* should expect educators to assess Criteria C and D together, which the further guidance document explains more clearly, according to our reviewer. Across Phases 1–3 for Criterion A, the same IB staff review suggested that the MYP *Guide* be more specific about “what conventions apply to these phases.” The IB staff review noted students with “prior exposure to language being taught” might be further along in learning oral language than they are in conventions, which the MYP *Guide* might be emphasizing too early for students with little to no knowledge of the target language.

¹³ Another educator found a problem across all phases for Criterion C, Strand i, in which the achievement level does not include the part of the objective stating “familiar,” “unfamiliar,” “social,” or “academic” situations. The educator stated that this problem was “not misalignment, but it is not clear why it has been left out of the achievement level descriptors.”

One educator identified an alignment problem for Phases 2 and 3 in Criterion C, Strand ii: the *MYP Guide* does not mention “limited variety of aspects” in descriptors for achievement level 7–8. The same educator found an alignment gap in Phase 3 where Criterion D, Strand i, was not aligned because the objective calls for a “range of vocabulary,” contrasting from achievement level 7–8’s expectation of a basic range of vocabulary. Also, the educator recommended that the *MYP Guide* should add “logical structure” and “cohesive devices add clarity to the message” for Criterion D, Strand ii, to Phase 3 and “to add to the message and to create a structured text” to Phases 3–6. For Phase 4, the educator found an alignment problem for Criterion C, Strand iii: achievement level 7–8 includes “opinions are supported by examples and illustrations,” so she suggested rephrasing the objectives for Phase 4–6 to include “with examples and illustrations in familiar and unfamiliar situations.” Last, IB staff found in Phase 6 that Criterion A’s command terms “do not reflect complexity in higher order thinking.”

Global Proficiency Table (Clarity)

Unlike the sections for previous analytical tasks, we do not include a table here because of the ease of interpreting our data. For the global proficiency table, all five reviewers rated the emergent (1 and 2) and capable phases (3 and 4) as Clear. All reviewers agreed that the first of the two proficient phases (5) was also Clear; one reviewer rated a lack of clarity for Phase 6. The one IB staff review that rated a lack of clarity did so because the Phase 5 section unpacks “social and academic situations” as “topics of personal interest and global significance,” but Phase 6 includes no such specification. The other IB staff review found all phases to be Clear, but questioned the meanings of “simple” and “short” in the oral and written assessment tasks.

Global Proficiency Table (Alignment)

In Table 3, we report the number of reviewers who rated the intersection of assessment criteria and a given phase as Aligned in the global proficiency table. Patterns emerged for Criteria C (communicate in response) and D (use language in spoken or written form) such that all reviewers rated Phases 2–5 as Aligned and almost all rated Phases 1 and 6 as Aligned. Four of 5 reviewers rated Criteria A (comprehension spoken/visual) and B (written/visual) as Aligned for Phases 3, 5, and 6. All reviewers rated Phase 4 as Aligned. At Phases 1 and 2, Criterion A was less consistently considered Aligned (3 of 5 reviewers) than was Criterion B (4 of 5 reviewers).

Though we did not see role-related differences in the quantitative data, the qualitative responses showed some association with whether a reviewer was an educator or IB staff. For example, both IB staff reviews recommended that the term “conventions” be made more explicit. One IB staff review also requested that command terms in Phase 6 should show a more pronounced progression from Phase 5 to reflect higher-order thinking skills such as analysis, interpretation, or evaluation.

Table 3. Number of Reviewers (N = 5) Rating Global Proficiency Tables as Aligned by Phase

Assessment criterion	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6
A. Comprehension spoken/visual	3	3	4	5	4	4
B. Comprehension written/visual	4	4	4	5	4	4
C. Communicate in response	4	5	5	5	5	4
D. Use language in spoken or written form	4	5	5	5	5	4

Two of the educators offered recommendations for Phases 1, 3, and 5. For Phase 1, one educator questioned whether the term “thorough” implied exactitude for Criteria A and B. The same educator wondered about a tension between using the terms “main ideas” or “basic ideas” within those Criteria. For Criterion C, the educator characterized the expectation of a Phase 1 learner to show an “excellent” sense of audience should instead be “some sense of audience.” In Criterion D, the educator suggested the word “basic” was missing before the word “language.” In Phase 3, the educator recommended that Criterion A, Strand i, moved the expectation of drawing conclusions to Phase 4. In Phase 5, another educator found misalignment in Criteria A and B. She noted that analysis is a more higher-order skill than recognition, so the expectation that a proficient communicator “recognize implied opinions and attitudes” alongside Strand iii assessment criteria that seeks learners who “engage thoroughly with the written/spoken/visual text by analyzing ideas, opinions and attitudes” demonstrates misalignment.

Language Acquisition Continuums (Clarity)

In Table 4, we report the number of reviewers who rated the language acquisition continuums as Clear for specific expectations and evidence across phases for 3-of-4 language continuums (all but writing, which all reviewers rated as Clear for both descriptions across all six phases). Though 4 of 5 reviewers rated viewing and interpreting as Clear for both descriptions and for all six phases, examining patterns for the data in this analytical task is less important than noting that 17 of the 18 ratings of Not Clear came from a single reviewer (94%). Still, three reviewers provided qualitative data, which dealt both with clarity and perceived level of challenge.

Table 4. Number of Reviewers (N = 5) Rating Phase-Specific Objectives as Clear by Phase

Continuum	Descriptions	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6
Listening and speaking	Specific expectations	5	5	4	5	5	5
	Evidence	5	5	4	4	5	5
Viewing and interpreting	Specific expectations	4	4	4	4	4	4
	Evidence	4	4	4	4	4	4
Reading comprehension	Specific expectations	5	5	5	5	5	5
	Evidence	5	5	4	5	3	5

Clarity. One IB staff review identified five issues that pertained to clarity. Regarding listening and speaking, the IB staff review wondered why these skills have been grouped together and seem to be accorded equal weight when listening is receptive and speaking is productive. Within that continuum, the IB staff review sought “a better way of phrasing ‘rehearsed and unrehearsed.’” The IB staff review expressed further confusion about why viewing and interpreting deserved its own continuum. The IB staff review acknowledged the value of the skill and stated it “should be an element of the curriculum,” but not “worthy of a continuum of its own.” The review characterized viewing and interpretation as a media skill, one in which interpretation led to subjectivity. Regarding reading comprehension, the IB staff review questioned why Phase 5 learners had to “list and explain the effect of the use of various stylistic devices or literary features in the text” but Phase 6 learners did not. Moreover, for both the reading comprehension and writing continuums, the IB staff review noted either an overlap or identical wording across phases. Last, one educator questioned the meaning of “breaks down” in the evidence component for Phase 5.

Perceived level of challenge. The IB staff review cited several instances in which the *MYP Guide* conveyed an expectation that the reviewer felt would impose an inappropriate level of challenge upon learners. Regarding listening and speaking, the IB staff review characterized the Phase 3 expectation that learners “understand and respond to a limited range of spoken texts” as “easier” than the corresponding Phase 2 demand. By contrast, the IB staff review found activities such as retelling a story too challenging for Phase 3 learners in both the listening and speaking and reading comprehension continuums.

Language Acquisition Continuums (Alignment)

Due to all five reviewers rating all comparisons for this analytical task as Aligned, we did not include a table to depict the alignment of the language acquisition continuums’ descriptions at achievement levels 7–8 by phase. Two reviewers—one of whom was an educator—offered minimal suggestions. The educator thought the *MYP Guide* was clear overall in this section, but that IB could enhance clarity in Phase 1 by replacing the term “basic-limited.” IB staff did not think the examples of evidence aligned well with the expectations.

Phase-Specific Assessment Task Guidelines (Appropriateness)

In Table 5, we report the number of reviewers who rated the phase-specific assessment task guidelines as Appropriate. Only one reviewer provided a rating of Not Appropriate for any phase other than Phase 3. Given the clustering of Not Appropriate ratings in Phase 3, we will report qualitative findings that discuss that phase first and then provide findings from the single reviewer who rated other phases as Not Appropriate.

Table 5. Number of Reviewers (N = 5) Rating Phase-Specific Assessment Task Guidelines as Appropriate

Criterion	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6
A. Comprehension spoken/visual	5	4	1	5	5	5
B. Comprehension written/visual	5	4	1	5	5	5
C. Communicate in response	5	5	4	5	5	5
D. Use language in spoken or written form	5	5	4	4	4	4

Regarding Phase 3, IB staff and two educators objected to the MYP *Guide* allowing learners to respond to the comprehension tasks in their mother tongues, noting “a huge conflict with the ePortfolios where ‘capable’ level learners must answer in the target language,” a conflict that could hinder students’ learning progressions. One educator recognized “students at their emergent level of learning the target language may need to use mother tongue when dealing with Criteri[a] [A and B, especially when they need to express personal experience and opinions or make personal response[s] to the text.” But based on her experience marking ePortfolios in 2016, the educator noticed, “Some schools abused or misapplied the claim: ‘All questions and answers may be in mother tongue, language of instruction or the target language.’” When marking portfolios, the educator found “whole assessment papers” for Tasks 1 and 2 that never “use[d] any target language.” The other educator relied on her experiences, which suggest Phase 2 learners as being capable of receiving questions for comprehension tasks only in the target language, though they may need to answer in their mother tongue or language of instruction. By Phase 3, the educator expected learners to be able to answer comprehension questions with spoken/written and visual text in the target language.

All three educators expressed concern about word counts in assessment tasks, though they did not agree on how those counts should change. One educator suggested a scale that would start with a 100–150 word range at Phase 1, increasing 50 words per phase and culminating in a 350–400 word range at Phase 6. Another educator concluded Phase 4–6 learners “are capable of writing more than is stated” in the MYP *Guide* and recommended a range with a ceiling of 600–800 words. The educator who sought to expand the ranges noted, “With more formal text types, such as essays or speeches, students struggle to meet the achievement level” within the existing word limits. The third educator had no issue with the MYP *Guide*’s general expectation, but noted the length would not be appropriate for Chinese. The educator suggested IB consider an alignment process to reduce confusion between the word length expectations for the text and writing tasks in the Chinese version of the MYP *Guide* and the *Guide to MYP eAssessment*.

Using Progressions to Place Students and Plan Instruction

The five reviewers responded to an exit interview item that asked about the MYP *Guide*’s effectiveness in helping place students and plan instruction. The group expressed consensus that the progressions are effective, though each reviewer stated one or more caveats. An educator described the guide as providing “generally clear information for teachers to classify students into different phases,” characterizing the assessment criteria, continuums, and evidence as “well-written and aligned with each other.” Another educator could see how the MYP progressions “work together” with DP expectations. One educator and one IB staff review each regarded the progressions as useful to make judgments about student placement and to inform instructional planning.

Despite the agreement, IB staff each noted one caveat, while the educators each presented several. One IB staff review did not believe that the “evidence” in the continuums aligns with the expectations and/or assessment criteria, a disconnect that might challenge educators. Another IB staff review noted the MYP *Guide* requires educators “to carefully consider each one of [the learning progressions] and know their purpose and integrate them” in placement, planning, and instruction. One educator noted the MYP *Guide*’s limited explicit consideration for differentiating among language needs of native students and requested additional support from IB in differentiating by ability in general. The same educator wondered why the MYP *Guide*

makes “no reference to mother tongue,” which some students want “to learn properly.” Instead of an absence, another educator criticized “an overemphasis on skills of visual comprehension,” which the reviewer felt should not “be equated with speaking and written skills to the extent that they currently are.”

Two educators described a disproportionately wide gap between Phases 3 and 4, which might be problematic for learners. This gap led one educator to characterize the progressions as “generally effective” but “not a smooth progression.” Another educator noted bigger gaps between Phases 2 and 3 and Phases 3 and 4 than the other phases, citing increased skill expectations in the global proficiency table as the cause. For example, Phase 3 expects learners to “speak and write in different ways for different purposes and audiences” before Phase 4 expects them to “communicate substantial information containing relevant and developed ideas and justified opinions on events, experiences and some concepts explored in class.” Similarly, in the assessment clarification for achievement level 7–8 of Criterion C, Phase 2 learners must respond in “basic structured exchanges.” Phase 3 learners must respond in “rehearsed and unrehearsed exchanges.” One reviewer asked IB to consider “how many students can really reach Phase 6?” Correspondingly, that educator questioned the MYP *Guide*’s claim that Phase 5 and 6 students could continue their target language studies as a Language A course in the DP. The educator felt that claim “is just too ideal” and that a more appropriate placement for Phase 5 and 6 students would be the Language B Higher Level (HL) course.

Phase III: Cross-Document Analysis

In Table 6, we report interrater reliability statistics for the reviews of the MYP *Guide*, overall and disaggregated by reviewer’s professional role (i.e., IB staff or MYP educator), and for the pairs of reviews for each of the international documents. For each interrater reliability analysis, pairwise agreement statistics ranged from 51% to 81% and reliability above chance agreement ranged from $\alpha = .00$ to $\alpha = .19$, both within the bounds of slight reliability (.00 to .20), according to Landis and Koch’s (1977) guidelines.

Table 6. Reviewers’ Interrater Reliability Statistics by Document

Document	Decisions	Average pairwise agreement (%)	Reliability above chance agreement (α)
MYP <i>Guide</i> ($N = 6$ reviewers)	222	68	.11
MYP <i>Guide</i> ($n = 3$ IB staff)	111	72	.15
MYP <i>Guide</i> ($n = 3$ educators)	111	67	.19
CEFR ($n = 2$ reviewers)	74	81 ^a	.00
ACTFL ($n = 2$ reviewers)	74	65 ^a	.15
HKEB ($n = 2$ reviewers)	74	51 ^a	.11

^a Analyses reflect pairwise agreement, not average pairwise agreement because there were only two reviewers for the given international document.

Note. Average pairwise agreement is the mean percentage agreement of each possible pair among the five reviewers. Reliability above chance agreement (i.e., Krippendorff’s alpha) shows the reviewers’ level of

agreement above and beyond the possibility of agreeing at random. All analyses included 37 cases, one for each of the analytical tasks. The number of decisions that raters made varied because they reflect the number of cases multiplied by the number of reviewers.

Comparing Documents

Overall, the *MYP Guide* and the CEFR provided the highest proportion of consensus ratings of supplying “enough” guidance. Five or more reviewers identified the *MYP Guide* as providing the right amount of guidance for 28 of 37 components (76%). The pair of CEFR reviewers agreed that the European document provided “enough” guidance for 29 components (78%). By contrast, reviewers of the ACTFL standards and HKEB standards agreed on “enough” as the rating for 21 components (57%) and 14 components (38%), respectively.

Table 7 reflects comparative ratings for the *MYP Guide* and the three international standards sets. The hue of the boxes ranges from white (*not at all*) to light blue (*too little*) to sea blue (*enough*) to dark blue (*too much*). EPIC coded one component of the *MYP Guide*—digitally mediated environments—as grey because its ratings varied from *enough* to *too little* to *not at all*, the only component to show such range. Ratings of the *MYP Guide* showed strengths and areas of improvement as follows, organized by the eight language acquisition domains that CASLS and EPIC found in the literature:

- **Interpersonal communication:** The *MYP Guide* and CEFR were particularly strong in this domain, whereas ACTFL and HKEB were relatively weak.
- **Defining phases of language acquisition:** For the most part, the *MYP Guide* and CEFR demonstrated strengths that ACTFL and HKEB did not. The *MYP Guide* appeared stronger than CEFR in focusing on communication acts, rather than syntactic or lexical accuracy.
- **Communicative competence:** The *MYP Guide* looked equally strong or comparatively stronger than the other documents for every competency except strategic.
- **Effective input, comprehension and interaction, and output:** The *MYP Guide* was as strong as the other documents in requiring or encouraging learners to analyze a variety of text types and in emphasizing receptive communication in which meaning is negotiated. Ratings for summative and formative assessment components of this domain suggest opportunities to improve the *MYP Guide*, with CEFR serving as a possible model for such improvement.
- **Intercultural competence:** Although ratings suggest an opportunity to improve in critical cultural awareness, the *MYP Guide* also seemed to be the strongest document overall for intercultural competence.
- **Contextualized communication:** All four documents demonstrated strengths in both components of this domain: language as cognitive and socially mediated.
- **Multimodality and multiliteracies:** *MYP Guide* ratings indicate a deficit in guidance on digitally mediated environments and face-to-face interactions.
- **Assessment:** The *MYP Guide* shows insufficient guidance on synchronous interactive writing (e.g., text messaging) and attending to errors. It should be noted, however, that ratings for all four documents on synchronous interactive writing tended toward *too little* or *not at all*.

Table 7. Comparative Ratings of the MYP Guide and Three International Documents

Domain	MYP Guide	CEFR	ACTFL	HKEB
<i>Interpersonal communication</i>				
Oral	Enough (5)	Enough (2)	Enough (2)	Too much / Enough
Written	Enough (6)	Enough (2)	Enough / Not at all	Enough / Too little
<i>Defining phases of language acquisition</i>				
Complexity increases	Enough (5)	Enough (2)	Enough (2)	Enough / Too little
Topics/content both vary and broaden	Enough (5)	Enough (2)	Enough / Too little	Enough / Too little
Reflects proficiency movement	Enough (6)	Enough (6)	Enough (2)	Enough (2)
Progression is logical, informed by expertise	Enough (5)	Enough (5)	Enough (2)	Enough (2)
Communication acts as focus	Enough (5)	Too much / Enough	Enough (2)	Enough / Too little
<i>Communicative competence</i>				
Grammatical	Enough (5)	Enough (2)	Enough / Too little	Enough (2)
Sociolinguistic	Enough (5)	Enough (2)	Enough (2)	Too much / Enough
Discourse	Enough (6)	Enough (2)	Enough (2)	Too much / Enough
Strategic	Enough / Too little	Enough (2)	Enough / Too little	Enough / Too little
<i>Effective input, comprehension and interaction, and output</i>				
Requires/encourages various text types	Enough (5)	Enough (2)	Enough (2)	Enough (2)
Oral communication between people	Enough (5)	Enough (2)	Enough (2)	Too much / Enough
Written communication between people	Enough (6)	Enough (2)	Enough / Too little	Enough / Too little
Interaction of reader and text	Enough (5)	Enough (2)	Enough (2)	Enough / Too little
Receptive communication with negotiated meaning	Enough (6)	Enough (2)	Enough (2)	Enough (2)
Summative assessment	Enough / Too little	Enough (2)	Enough / Too little	Enough (2)
Formative assessment	Enough / Too little	Enough (2)	Enough / Too little	Enough / Too little

Domain	MYP Guide	CEFR	ACTFL	HKEB
<i>Intercultural competence</i>				
Interpreting and relating	Enough (5)	Enough / Too little	Enough / Too little	Enough / Too little
Discovery and interaction	Enough (5)	Enough / Too little	Too little (2)	Too little (2)
Critical cultural awareness	Enough / Too little	Enough / Too little	Enough (2)	Too little (2)
<i>Contextualized communication</i>				
Language is cognitive	Enough (5)	Enough (2)	Enough (2)	Enough (2)
Language is socially mediated	Enough (5)	Enough (2)	Enough (2)	Enough (2)
<i>Multimodality and multiliteracies</i>				
Digitally mediated environments	Conflicted	Enough (2)	Too little / Not at all	Too little (2)
Face-to-face interactions	Enough / Too little	Enough (2)	Enough / Too little	Enough (2)
Classroom interactions	Enough (6)	Enough (2)	Enough / Too little	Enough (2)
<i>Assessment</i>				
Presentational speech	Enough (5)	Too much / Enough	Enough (2)	Too much / Enough
Synchronous interactive writing	Too little / Not at all	Enough / Too little	Too little / Not at all	Not at all (2)
Listening comprehension	Enough (5)	Enough (2)	Too little (2)	Enough (2)
Text-based comprehension	Enough (5)	Enough (2)	Enough (2)	Enough / Too little
Unimodal demonstrations	Enough (5)	Enough / Too little	Enough (2)	Enough / Too little
Multimodal demonstrations	Enough (5)	Enough (2)	Too much / Enough	Too much / Enough
Realistic	Enough (6)	Enough (2)	Enough (2)	Enough (2)
Performance-based	Enough (6)	Enough (2)	Enough (2)	Too much / Enough
Cognitively complex	Enough (5)	Enough (2)	Enough (2)	Enough (2)
Meaningful communication	Enough (6)	Enough (2)	Enough (2)	Enough (2)
Attending to errors	Enough / Too little	Enough (2)	Enough (2)	Enough (2)

Unique Strengths

First we report on the strengths of the *MYP Guide*. In subsequent sections, we report on the strengths that pairs of reviewers found for the CEFR, ACTFL, and HKEB, respectively.

MYP Guide. We report findings about the *MYP Guide's* unique strengths in descending order of the frequency with which reviewers indicated a strength.

5 of 6: All but one reviewer recognized the *MYP Guide* for presenting an authentic, holistic, conceptual, and inquiry-based approach. One educator appreciated the *MYP Guide* for allowing students to learn through projects and interactions/expression of ideas with peers and teachers. Another educator credited the *MYP Guide's* “lack of prescriptive topics and text types and the wide range of contexts to choose from,” which facilitates greater lesson-plan creativity. The third educator enjoyed the encouragement for students to apply learning in authentic situations that involve appreciation of one’s own and other cultures.

4 of 6: Two IB staff and two educators each appreciated the *MYP Guide's* four communicative skills (speaking, listening, reading, and writing). One IB staff review praised the *MYP Guide* for weighting the skills equally. Another IB staff review was pleased that the *MYP Guide* allows for development in those areas, but was troubled that Assessment Criteria C and D do not address all four skills. Two educators appreciated a similar aspect, but presented opposite viewpoints about it. One educator praised the skills for not following a linear progression, therefore, enabling a student to excel in one area, while struggling with another. The other educator suggested that students “must develop the four communicative [skills] . . . evenly to progress through the phases,” and considered the ACTFL standards superior in this area. The second educator cited the case of a heritage speaker: for example, one might be fluent in oral Chinese but not know how to write Chinese characters. The educator noted the inappropriateness of assessing the student at novice range for all skills if that range applies only to one competency.

3 of 6: The reviewers who were educators all welcomed the *MYP Guide's* integrative approach for facilitating vertical planning and reflection on students’ needs and goals, according to one educator. Another educator appreciated the emphasis on developing sociolinguistic competency¹⁴ via global contexts, though she noted that concept-based instruction might be more effective once students have completed the beginning levels of language acquisition. The third educator enjoyed how key concepts, global contexts, and other facets of the *MYP Guide* help students “solve real-world problems.”

3 of 6: Two IB staff praised the *MYP Guide* for embedding the IB philosophy in its objectives and aims, particularly for being culturally inclusive of all languages. However, one staff review noted that respect and understanding for linguistic and/or cultural diversity are “not always apparent in the phase-level descriptors.” One educator characterized the phases overall as well designed, but noted some gaps. The teacher sought clearer guidelines to assist schools and teachers on

MYP Guide’s Unique Strengths

- Authentic, holistic, conceptual, and inquiry-based approach
- Integration of speaking, listening, reading, and writing
- Facilitation of vertical planning
- Reflection on students’ needs and goals

¹⁴ Typically, this term appears in second-language acquisition literature as sociolinguistic competence, but we use the term sociolinguistic competency as did Siegal (1996), though in our instance we do so to maintain our distinction between competence and competency.

how to promote students from one phase to another, noting that variable interpretations would be problematic with the rise of eAssessment. The educator requested more examples to generate greater interpretive consensus about phase-to-phase gaps, especially between Phases 3 and 4.

1 of 6: One IB staff review appreciated that the *Guide* “is an actual curricular GUIDE, not simply an instrument for evaluation (although that is a large component of it)” because the document “covers everything from philosophy to programme requirements to student placement in levels and then level descriptors.”

CEFR. Responses to the CEFR included nondescript praise for its clear connection to the MYP *Guide*, which the educator reviewer identified as having “the same ‘goals’ in terms of student achievement and proficiency” despite using different vocabulary. Both documents sought to help students understand “language, culture, and the society where the language is spoken in order to be able to communicate and express their ideas, feelings or points of view, and be understood although having some few or minor errors.” The IB staff review felt that CEFR is “a good tool that can be adapted” because it shares the MYP *Guide*’s focus on “the rich and diverse cultural heritages” of learners.

CEFR

Shares goals with MYP *Guide* but uses different verbiage

ACTFL. ACTFL reviewers agreed on two instances of its unique strengths and then each identified strengths that the other did not. Both reviewers concluded that ACTFL provided clearer description and exemplification than the MYP *Guide*. Although the educator noted that the progression of syntactic, grammatical, and lexical accuracy was stronger in the ACTFL than the MYP *Guide*, she suggested that the ACTFL clarify terms such as “advanced-level control of grammar and syntax.” Also, both reviewers appreciated ACTFL’s progressions from performance to proficiency. The educator found the progression to be encouraging and flexible, especially for students whose skills might regress. The IB staff review believed that the three levels¹⁵ in ACTFL’s *Standards for Foreign Language Learning in the 21st Century* might be more useful than MYP’s six, particularly for

ACTFL

Clearer descriptions and exemplification than MYP *Guide*

MYP schools that only include Years 9 and 10. Additionally, the educator praised ACTFL for emphasizing (a) use of the target language beyond what she observed in the MYP *Guide*, (b) listening comprehension with speaking in the interpersonal performance descriptors, and (c) consistent use of terms such as “appropriateness” and “accuracy,” which are not interchangeable, but used that way in the MYP document. IB staff praised ACTFL’s concise design for being more useful and less overwhelming, especially for newer IB practitioners. That being said, IB staff recognized that the MYP *Guide* has “a larger scope” and is intended to connect to the rest of the IB programme, so it remains unclear how to “keep the intent of the MYP *Guide* without sacrificing context to make it more concise or user-friendly.”

HKEB. HKEB reviewers produced three suggestions for how the MYP *Guide* could be improved, overlapping on two. Both reviewers noted the HKEB’s focus on enjoyment and lifelong learning. The educator specifically called for the MYP *Guide* to incorporate development of students’

¹⁵ The *Standards for Foreign Language Learning in the 21st Century*, which we reviewed in this study, references only three levels of proficiency (i.e., novice, intermediate, advanced). However, *Proficiency Guidelines 2012*, another important ACTFL document, includes an additional two levels of language proficiency (i.e., superior and distinguished).

values and positive attitudes alongside their knowledge and skills, noting that the Approaches to Learning do not explicitly focus on dispositional elements. Furthermore, IB staff appreciated the HKEB's twin focus on formative and summative assessment. The reviewers, however, departed in one area: the HKEB's provision of "lots of strategies and examples," which the educator believed would reduce confusion:

HKEB

Greater focus than MYP *Guide* on student enjoyment, lifelong learning, development of students' values and positive attitudes

Teachers will definitely have . . . clearer concepts or ideas on how to teach and apply the curriculum in everyday teaching. With sufficient examples or teaching strategies, I am sure teachers will not get confused easily when they are either setting assessment papers or putting students to the right phase. (Reviewer 3)

IB staff characterized the same facet differently, noting how the HKEB "prescribes" grammar, language skills, and strategies, using a decidedly different tone about the nature of providing guidance. Meanwhile, the educator appreciated the HKEB's combination of "compulsory and elective parts" that enable "students to choose according to their own preference." HKEB's approach allows teachers to cater to varied learning needs, add curricular variety, and motivate students by appealing to their interests. The educator recommended that MYP adopt a similar approach.

Strengths

- Meaningful communication
- Authentic contexts
- Intercultural competence

Phase IV: Progression Analysis

In Table 8 we outline how each IB programme describes language acquisition and addresses the critical influences that we identified in our literature review (e.g., input, negotiation of meaning, communicative context, and intercultural competence). We organized Table 8 into nine categories:

1. Conceptualization of effective language teaching and learning
2. Intercultural competence
3. Organizing features of learning
4. L1/L2 integration
5. Phases of communication
6. Areas of communication
7. Communication continuums
8. Communicative processes
9. Approaches to assessment

Our literature review demonstrated the importance of considering the following three categories: (a) conceptualization of effective language teaching and learning, (b) intercultural competence, and (c) approaches to assessment.

We included the following additional categories into these analyses to address the possible impact of transitioning from concept-based education in the PYP and the MYP to topic-based

Areas to Consider

Cross-programme variation in

- Terminology
- Framework
- Assessment approaches

education in the DP, also highlighting when second-language acquisition receives treatment as a subject area separate from L1 acquisition in the IB continuum: the organizing features of language learning and L1/L2 integration. After considering the framework for language acquisition articulated in the *MYP Guide*, we also added into these analyses phases of communication, areas of communication, communication continuums, and communicative processes.

Table 8. *Progression Analysis of Language Acquisition in the PYP, the MYP, and the DP*

Category	Primary Years Programme	Middle Years Programme	Diploma Programme
Conceptualization of effective language teaching and learning	All three programmes' documentation demonstrates that <ul style="list-style-type: none"> language is an integral component of communication, and a social act mediated by context, environment, relationships, and self meaning is interpreted from input and is more important than knowledge of vocabulary, grammar, and syntax interaction is emphasized in communication acts involving reception output does not simply prove that language acquisition has occurred; it is process that is integral to language acquisition 		
Intercultural competence	All three programmes' documentation demonstrates that the development of intercultural competence is both the goal and the purpose of language learning		
Organizing features of learning	Scope and sequence involves following key conceptual understandings: <ul style="list-style-type: none"> Form Function Causation Change Connection Perspective Responsibility Reflection 	Curriculum involves following key conceptual understandings: <ul style="list-style-type: none"> Communication Connections Creativity Culture Additional concepts are differentiated by phase: accent, audience, context, conventions, form, function, meaning, message, patterns, purpose, structure, word choice, empathy, idiom, point of view, argument, bias, context, inference, stylistic choices, theme, and voice.	Conceptual approach not mentioned in <i>DP Language B Guide</i> . Instead, learners acquire language via exploration of 3 core topics (communication and media, global issues, and social relationships) and 2-of-5 possible optional topics (cultural diversity, customs and traditions, health, leisure, and science and technology). Within each of the 5 topics selected, learners must study 2 related aspects (e.g., if studying health, learners may study eating disorders and global health pandemics).
L1/L2 Integration	L1 and L2 development are addressed with the same documents and set of principles; language is transdisciplinary.	Clear between L1 and L2 acquisition. Language is important across disciplines, but the <i>MYP Guide</i> focuses on the L2 context.	Distinction between L1 and L2 acquisition with the possibility of identifying two L1s for a given learner in lieu of one L2; language (L1 and L2) is featured as a way of knowing across disciplines in the Theory of Knowledge course.

Phases of communication	5: specified learning outcomes ¹⁶ at each phase, independent of age.	6: specified learning outcomes at each phase, independent of age.	Not mentioned in <i>DP Language Acquisition Guide</i> .
Areas of communication ¹⁷	<ul style="list-style-type: none"> • Oral • Written • Visual 	<ul style="list-style-type: none"> • Oral • Written • Visual 	Not explicitly mentioned in <i>DP Language B Guide</i> . However, assessment tasks do involve oral, written, and visual components.
Communication continuums ¹⁸	Define learning targets as: <ul style="list-style-type: none"> • listening & speaking • viewing & presenting • reading • writing 	Define learning targets as: <ul style="list-style-type: none"> • listening & speaking • viewing & interpreting • reading comprehension • writing 	Not mentioned; instead, the <i>DP Language B Guide</i> outlines that language acquisition occurs through the development of receptive, productive, and interactive skills and competencies.
Communicative processes	None listed; assessments are part of transdisciplinary units and focused on revealing learners' progression through the Learning Continuums.	Form assessment criteria: <ul style="list-style-type: none"> • Comprehending spoken & visual text • Comprehending written & visual text • Communicating in response to spoken and/or written and/or visual text • Using language in spoken and/or written form Each strand within each criterion must be assessed summatively twice in a given school year.	Competencies and skills defined as receptive, productive, and interactive. Assessment tasks include: <ul style="list-style-type: none"> • text handling/reading comprehension • written productive skills (two assessment criteria) • receptive and written productive skills (three assessment criteria) • two interactive oral activities each with two assessment criteria (with peers and with the teacher in which a presentational component is mandated)
Assessment approach	All three programmes' documentation demonstrates that assessment approaches involves both process (formative) and products (summative) of learning.		
Assessment approach: Rubric/markscheme development	Teachers develop materials that can include rubrics.	IB-provided rubrics are aligned with communicative processes and differentiated by phase.	IB-provided rubrics/markschemes differentiated by course (Standard Level [SL] or Higher Level [HL]).
Assessment approach: Mentions of tasks	Possible tasks are mentioned.	Possible assessment tasks are articulated.	Assessment task design carefully articulated.

¹⁶ These learning outcomes are similar in PYP and MYP. For example, picking out main ideas shows up in Phase 3 of the Listening and Speaking Continuum in both programmes. However, each guide's wording of the learning outcomes is distinct.

¹⁷ The areas of communication are referred to as "strands" in PYP literature. "Strands" refers to assessment descriptors for MYP and DP.

¹⁸ The "Communication Continuums" are referred to as "Learning Continuums" in PYP.

Assessment approach: Internal vs. external assessments	Not applicable	Internal assessments developed by teachers and are an MYP assessment cornerstone; externally marked e-portfolio is encouraged at programme’s end, but not required.	Internal assessments that are externally moderated and external evaluation at the program’s end are required.
Assessment approach: Teacher development	Teachers develop tasks to address the Learning Continuums.	Teachers do not develop the e-portfolio tasks.	Teachers do not develop externally moderated tasks but can influence text choice in written assignments.

The following sections detail our comparative analysis of each IB programme’s treatment of language acquisition. We organized the analysis by the categories featured in Table 8. We highlight points of alignment and misalignment to establish how well the progression of language development described in the MYP *Guide* links with the PYP and the DP. To facilitate readers’ examinations of those of our comparisons that revealed misalignments, we provide the relevant Table 8 topic immediately preceding the corresponding comparative analysis.

Conceptualization of Effective Language Learning and Teaching

Contemporary second-language acquisition research emphasizes input, interaction and negotiation of meaning, and output as foundational components of language acquisition. Understandings of the social framework, purpose, and context of communication are essential to the success of communication, be it unidirectional (interpretive or presentational) or bidirectional or multidirectional (interpersonal). Although a lower-level skill such as grammatical competency might be important to clear communication of meaning, higher-level skills that more directly involve meaning making (e.g., sociolinguistic competency, discourse competency, and strategic competency) should be the overwhelming focus in courses designed to support target language acquisition. That each of the IB guides we reviewed for this analysis reflects these foundational principles for language learning is undeniable; all guides explicitly value function over form and meaning over accuracy. Indeed, the overwhelming goal of each programme is to develop L2 users who are able to participate in meaningful communication.

By extension, each IB guide that we reviewed for this analysis valorizes the development of language proficiency in learners more so than their acquisition of discrete vocabulary and grammar points. For example, in *PYP Language Scope and Sequence*, language learning and teaching are described as “social acts, dependent on relationships with others, with context, with the environment, with the world, and with the self” (p. 10). The MYP *Guide* echoes this sentiment, citing Savignon (1983): “Learning to speak another’s language means taking one’s place in the human community. . . . Language is . . . our most important link to the world around us. . . . It is people interacting with people” (p. 4). The DP again reinforces the communicative purpose of language learning, describing the Language *ab initio* and Language B courses as being geared toward necessary skills and intercultural understanding that enable students to communicate successfully in environments where their target language is spoken. The foundational principles of language acquisition expressed in each guide clearly align with one another.

Intercultural Competence

Though none of the guides specifically discuss Byram's (1997) notion of intercultural competence,¹⁹ its presence is undeniable. The purpose of language learning as discussed in *PYP Language Scope and Sequence* provides a representative example:

Exposure to and experience with languages, with all their richness and diversity, creates an inquisitiveness about life and learning, and a confidence about creating new social interactions. Language provides a vehicle for learners to engage with the world and, in an IB World School, to relate to, and accept, responsibility for the mission of the IB to “help to create a better and more peaceful world.” (p. 10)

This belief that language learning can lead to the development of more empathetic, understanding individuals echoes across programmes. For example, the MYP “requires the study of at least two languages to support students in understanding their own cultures and the cultures of others” (p. 3, *MYP Guide*), and the *DP Language B Guide* articulates that language acquisition courses “are designed to provide students with the necessary skills and intercultural understanding to enable them to communicate successfully . . . beyond the confines of the classroom” (p. 4). Across programmes, language is contextualized as a vehicle that students must actuate as they navigate distinct peoples and perspectives. It is this process of familiarization that increases intercultural awareness and understanding. This alignment is essential to the mission of the IB, allowing educators across programmes to use a similar lens when emphasizing to stakeholders the critical importance of language learning.²⁰

Organizing Features of Learning

As Spada (2005) discusses, the successful execution of communication acts is influenced by any number of factors including power dynamics among interlocutors, the social context in which communication is taking place, the cultural backgrounds and perceptions of interlocutors, and the linguistic and grammatical competencies of interlocutors. Hyland (2007) contributes to this conversation by extending the charge that educators prepare learners to communicate within all social contexts and for a variety of purposes, not just within those which are typical to the academic sphere. Indeed, understanding and negotiating with language within a variety of contexts is requisite of any language learner desiring to fully develop his or her proficiency.

In alignment with this requirement, each IB programme emphasizes that language communication be meaningful and contextualized (see Table 9). However, their approaches to achieving this type of communication are distinct. The PYP and the MYP emphasize concept-based learning to provide a lens for the contextualization of communication acts; the *DP Language B Guide* employs a topic-based approach. It is worth noting, however, that Approaches to Learning (ATL) documents published after the *DP Language B Guide* in the Diploma Programme emphasize that promotion of

¹⁹As we articulated in the literature review, intercultural competence involves the degree to which an individual can be confronted with a previously foreign cultural artifact and not regard it as less worthy than the artifacts from his or her own culture.

²⁰ The IB Programme Standards and Practices illuminate this alignment well: standard A4 emphasizes that an IB school “develop and promote international-mindedness and all attributes of the IB Learner Profile across the school community,” and standard A7 emphasizes placing “importance on language learning including mother tongue, host country, and other languages” (p. 3).

conceptual understandings should drive teaching. Additionally, the DP unit planners published after the guide now include areas for educators to articulate content, skills, and concepts.

Table 9. *Organizing Features of Learning Across the IB Continuum*

PYP	MYP	DP
<p>Scope and sequence involves following key conceptual understandings:</p> <ul style="list-style-type: none"> • Form • Function • Causation • Change • Connection • Perspective • Responsibility • Reflection 	<p>Curriculum involves following key conceptual understandings:</p> <ul style="list-style-type: none"> • Communication • Connections • Creativity • Culture <p>Additional concepts are differentiated by phase: accent, audience, context, conventions, form, function, meaning, message, patterns, purpose, structure, word choice, empathy, idiom, point of view, argument, bias, context, inference, stylistic choices, theme, and voice.</p>	<p>Conceptual approach not mentioned in <i>DP Language B Guide</i>. Instead, learners acquire language via exploration of 3 core topics (communication and media, global issues, and social relationships) and 2-of-5 possible optional topics (cultural diversity, customs and traditions, health, leisure, and science and technology). Within each of the 5 topics selected, learners must study 2 related aspects (e.g., if studying health, learners may study eating disorders and global health pandemics).</p>

Contemporary language acquisition literature is mute regarding the relative impact of concepts and topics in language acquisition for learners.²¹ It does, however, emphasize the importance of context to communication and negotiation of meaning, something with which all IB programmes align well.

L1/L2 Integration

As Table 10 illustrates, the PYP is dissimilar from the MYP and the DP in that programme documents articulate the language learning expectations of L1 and L2 learners simultaneously. Although learners are certainly not expected to be at the same phase of language acquisition in an L1 as in an L2, they are expected to move through those phases similarly in both contexts. Conversely, the MYP and the DP distinguish L1 from L2 acquisition, not merely within the differentiation of learning expectations, but also in the treatment of the two as distinct subject areas. If a DP learner progresses far enough in the L2 that his or her proficiency is similar to that which he or she possesses in the L1, the learner may opt to study two L1s.

²¹ Though this assertion is true in the field of second-language acquisition, the concept-based approach features prominently in various IB publications as it emphasizes the cross-disciplinary, enduring nature of concept-based learning, which it esteems more than topic-based learning.

Table 10. *Phases of Language Acquisition Across the IB Continuum*

PYP	MYP	DP
L1 and L2 development are addressed with the same documents and set of principles; language is transdisciplinary.	Clear between L1 and L2 acquisition. Language is important across disciplines, but the MYP <i>Guide</i> focuses on the L2 context.	Distinction between L1 and L2 acquisition with the possibility of identifying two L1s for a given learner in lieu of one L2; language (L1 and L2) is featured as a way of knowing across disciplines in the Theory of Knowledge course.

Phases of Language Acquisition

The phases of language acquisition are specifically mentioned in both the PYP and the MYP documents, though the PYP discusses only five phases (perhaps because progression to the sixth stage is unexpected in the primary years), and the MYP discusses six. Still, IB’s defined learning outcomes for each of the common phases align well. Such alignment is not directly evident upon consideration of the DP guide. In the PYP and the MYP, phases and their associated learning outcomes indicate the proficiency levels of learners and should inform their class placement and the selection of differentiation strategies within the classroom. Although appropriate placement of learners into DP courses (i.e., Language B *ab initio*, Language B SL, or Language B HL) is discussed in the *DP Language B Guide* (see p. 6), discussion of phases is absent. Such an absence is particularly noticeable upon consideration of the articulation of connections between MYP and DP provided in the *DP Language B Guide* (p. 7). On a practical level, the inclusion of phases in the DP Language B Guide could be beneficial for establishing common language among educators working in schools along the IB continuum, especially when they engage in vertical articulation efforts. The possible continuum pathway outlined in the *MYP Guide* (p. 6) would be a useful document to support such an endeavor. See Table 11 for a summary.

Table 11. *L1/L2 Integration Across the IB Continuum*

PYP	MYP	DP
5: specified learning outcomes ²² at each phase, which is independent of age	6: specified learning outcomes at each phase, which is independent of age	Not mentioned in <i>DP Language Acquisition Guide</i>

Language Frameworks—Areas of Communication

PYP documentation makes specific mention under the alternative heading “strands of communication” to the areas of communication (oral, visual, and written) that the *MYP Guide* defines (see Table 12). Conversely, the discussion of areas of communication is largely absent from the *DP Language B Guide*, though the assessment tasks certainly involve visual interpretation (namely Paper 1, the Written Assignment, and the Individual Oral), oral / aural tasks (the Individual Oral and the Interactive Oral), and written tasks (namely Paper 2 and the Written Assignment).

²² These learning outcomes are similar in PYP and MYP. For example, picking out main ideas shows up in Phase 3 of the Listening and Speaking Continuum in both programmes. However, each guide’s wording of the learning outcomes is distinct.

Table 12. Areas of Communication Across the IB Continuum

PYP	MYP	DP
<ul style="list-style-type: none"> • Oral • Written • Visual 	<ul style="list-style-type: none"> • Oral • Written • Visual 	Not explicitly mentioned in <i>DP Language B Guide</i> ; assessment tasks involve oral, written, and visual components.

Language Frameworks—Communication Continuums

Upon consideration of the communication continuums defined in the *MYP Guide*, we see clear alignment between the MYP and the PYP (though the PYP refers to the continuums as “learning continuums;” see Table 13). Each programme includes the following developmental continuums: Listening and Speaking, Reading Comprehension,²³ and Writing. The single discrepancy is that the PYP involves a Viewing and Presenting continuum, a logical, reciprocal continuum that directly addresses presentational speech. Instead, the MYP incorporates a Viewing and Interpreting²⁴ continuum, which is nonreciprocal and does not specifically mention presentational speech acts. Neither guide specifically addresses synchronous interactive writing nor does either guide explain how these continuums materialized.

IB’s brief explanation of the continuums in the relevant guides seems to connect them most to the modes of communication as CEFR and ACTFL define them. Still, the similarities do not provide for an easy, one-to-one correspondence, a point that the relative absence of presentational speech acts in the *MYP Guide* illustrates. However, this absence is not evident in the *DP Language B Guide*: the expectation that learners’ receptive, productive, and interactive skills and competencies develop is emphasized (see p. 6; pp. 20–21). These skills and competences align directly with ACTFL’s interpretive mode (Interpretive Reading and Interpretive Listening), presentational mode (Presentational Speaking and Presentational Writing), and interpersonal mode. They also align with CEFR’s types of communication: Listening, Reading, Spoken Interaction, Spoken Production, and Writing. It is likely that this alignment is beneficial to practitioners dealing with the widely proliferated ACTFL and CEFR standard sets. By extension, it is possible that MYP practitioners and PYP practitioners would benefit from a similar degree of alignment.

Table 13. Communication Continuums Across the IB Continuum

PYP	MYP	DP
Define learning targets as <ul style="list-style-type: none"> • listening & speaking • viewing & presenting • reading • writing 	Define learning targets as <ul style="list-style-type: none"> • listening & speaking • viewing & interpreting • reading comprehension • writing 	Not mentioned; instead, the <i>DP Language B Guide</i> outlines that language acquisition occurs through the development of receptive, productive, and interactive skills and competencies.

²³ This continuum is simply called “Reading” in the PYP.

²⁴ It is worth noting that viewing and presenting are reciprocal acts as are speaking and listening. It is unclear why the MYP employed the reciprocal acts of speaking and listening on one continuum but did not employ reciprocal acts to define the Viewing and Interpreting continuum.

Language Frameworks—Communicative Processes

Misalignment among programmes in the IB continuum is most obvious when considering the communicative processes. As discussed in the literature review, the communicative processes in the MYP *Guide* (Assessment Criterion A: Comprehending spoken and visual text, Assessment Criterion B: Comprehending written and visual text, Assessment Criterion C: Communicating in response to spoken and/or written and/or visual text, and Assessment Criterion D: Using language in spoken and/or written form) are the criteria upon which learners are evaluated. IB-created rubrics inform and guide those evaluations.

Although the PYP is very similar to the MYP in its treatment of developmental continuums, it diverges from the MYP by employing no assessment criteria/communicative processes. Learning targets link to the PYP continuums, but no assessment criteria materialize from the disentanglement and combination of those continuums as in the MYP, and to some extent the DP. Because the thought process behind the materialization of these assessment criteria is unclear, it is difficult to assess which programme offers the superior articulation of learner expectations.

Given that both the MYP and DP outline assessment criteria, it is useful to compare these criteria to identify possible points of alignment and misalignment. Each of the following comparisons uses DP assessment tasks as a benchmark against the most relevant MYP assessment criteria:

- Paper 1 in the DP, a written text handling assessment, is notably distinct from MYP's Criterion B, which addresses reading comprehension. Paper 1 involves no personal response to texts being read and is graded with a markscheme in lieu of a rubric. In this sense, MYP Criterion B conforms more with the principles of authentic assessment (see Wiggins 1989; 2011) than Paper 1 does. The MYP criterion requires learners to use written texts to perform authentic and meaningful communication tasks. Conversely, Paper 1 assesses comprehension among DP learners without using authentic communication contexts.
- Paper 2 in the DP, a presentational writing task, resembles MYP Criterion D. Both tasks are evaluated with a rubric inclusive of descriptors that address form, content, and grammatical accuracy. However, Paper 2 diverges by allowing only for written responses instead of written and/or spoken responses. Though this MYP provision provides a mechanism for increased learner autonomy, it does also give learners a way to avoid either presentational writing or presentational speaking when executing assessment tasks.
- The Written Assignment in the DP, a text-based presentational writing task, combines the expectations set forth in MYP Criteria A (reading comprehension), C (interaction in response to a text), and D (presentational writing or speaking). Not only are DP learners evaluated for their writing skills (i.e., grammatical correctness, linguistic variety, and implementation of cohesive and rhetorical devices), but also for their ability to understand and incorporate the targeted text (HL) or texts (SL). Furthermore, DP learners must justify their choice of topic, text type, and target audience on the Written Assignment and must engage in a personal response to the text(s) at hand when engaging with MYP Criterion A. Undoubtedly, the Written Assignment aligns more with principles of authentic assessment as they relate to reading comprehension than does Paper 1.
- The Interactive Oral Activity in the DP, in which learners speak with each other about a topic, is highly aligned with MYP Criterion C. Both assessments evaluate interactive skills, as well as accuracy of language. However, presentational tasks can address MYP Criterion C; the Interactive Oral Activity must have interactive components.

- The Individual Oral in the DP, a spontaneous presentation in response to a photo stimulus with a follow-up interview between the teacher and the student in question, combines the expectations of MYP Criteria C and D well. The clear alignment between the tasks stems from their shared incorporation of response to a visual text.

In sum, though MYP and DP assessment activities align to a degree, this alignment is not direct. Evaluating either the MYP or DP approach as more effective than the other is entirely subjective: the MYP requires more authentic reading tasks of its learners; the DP better protects presentational speaking. Still, the degree to which they do not align may prove confusing to learners who move along the continuum and to practitioners teaching in multiple programmes.

Table 14. *Communicative Processes Across the IB Continuum*

PYP	MYP	DP
None listed; assessments are part of transdisciplinary units and focused on revealing learners' progression through the Learning Continuums	<p>Form assessment criteria:</p> <ul style="list-style-type: none"> • Comprehending spoken & visual text • Comprehending written & visual text • Communicating in response to spoken and/or written and/or visual text • Using language in spoken and/or written form <p>Each strand within each criterion must be assessed summatively twice in a given school year.</p>	<p>Competencies and skills defined as receptive, productive, and interactive. Assessment tasks include</p> <ul style="list-style-type: none"> • Text handling/reading comprehension • Written productive skills (two assessment criteria) • Receptive and written productive skills (three assessment criteria) • Two interactive oral activities with two assessment criteria each (one with peers and one with the teacher in which a presentational component is mandated)

Approach to Assessment

The previous discussion regarding the communicative processes points to the different approaches to assessment that the PYP, MYP, and DP employ. We highlight assessment requirements within each programme as illustrations of programmatic differences:

- In the PYP, assessment occurs within transdisciplinary units and is used to determine learners' development along four learning continuums. Practitioners use the *PYP Language Scope and Sequence* document to inform the creation of evaluative mechanisms such as rubrics and assessment tasks. PYP requires no external evaluation or moderation of learners.
- In the MYP, the language continuums are disentangled and combined, yielding to create the four communicative processes/assessment criteria. Each assessment criterion involves strands that serve as descriptors within the associated rubrics. IB provided these rubrics to practitioners for each phase of language acquisition. Within a school year, language acquisition teachers must evaluate each student on each strand for each criterion twice using summative assessments that teachers in the schools develop. Finally, an externally evaluated e-portfolio is encouraged at the end of MYP Year 5, but it is not yet required.

- In the DP, all identified summative assessment tasks other than the internal assessments (the Individual Oral Activity and the Interactive Oral Activity) are externally evaluated. IB requests a sample of each class’s Individual Oral Activities and trains other IB practitioners to moderate these samples externally to validate the instructor’s evaluation of learners and their work. IB may change instructors’ scores based on the moderation process. It is important to note that practitioners provide their learners with some input on the Written Assignment, even though they do not assess it themselves. It is similarly important to note that practitioners only report to IB each student’s highest score among three Interactive Oral Activities.

Although these distinctions (summarized in Table 15) are broad and may complicate understanding with specific programme goals and expectations of practitioners and students across the IB continuum, the foundational principles of assessment espoused within each programme align well. Each programme discusses that assessment should be formative (focused on providing learners and educators with information regarding the learners’ current level of performance) and summative (focused on providing “an overview of previous learning and . . . concerned with measuring student achievement [*DP Language B Guide*, p. 24]). Similarly, learners demonstrating understanding of the communicative context within a given assessment is essential to success, and all criterion-referenced assessments focus on meaningful communication.²⁵

Table 15. *Approach to Assessment Across the IB Continuum*

	PYP	MYP	DP
Rubric/ markscheme development	Teachers develop materials that can include rubrics.	IB-provided rubrics are aligned with communicative processes and differentiated by phase.	IB-provided rubrics/ markschemes differentiated by course (SL or HL).
Mentions of tasks	Possible tasks are mentioned.	Possible assessment tasks are articulated.	Assessment task design carefully articulated.
Internal vs. external assessments	Not applicable	Internal assessments developed by teachers and are an MYP assessment cornerstone; externally marked e-portfolio is encouraged at programme’s end, but not required.	Internal assessments that are externally moderated and external evaluation at the program’s end are required.
Teacher development of assessments	Teachers develop tasks to address the Learning Continuums.	Teachers do not develop the e-portfolio tasks.	Teachers do not develop externally moderated tasks but can have influence in choice of source text in the written assignment.

²⁵ Other than Paper 1 in the DP, all assessments across programmes should involve some degree of criterion-referenced evaluation.

Overall Progression Findings

Summatively, the *MYP Guide* illustrates foundational principles of language acquisition and assessment that link very well with the PYP and the DP. For example, ideal language learning in all programmes involves meaningful communication in a variety of authentic contexts. Furthermore, language acquisition is valued for its ability to improve the intercultural competence of learners. The differences among programmes are noted in the approach to implementing said principles. These differences include the terminology used, the framework provided for language acquisition, and the approach to assessment within each programme. With respect to the terminology used, we have highlighted that the language continuums in the MYP are referred to as learning continuums in the PYP and that “strand” within the MYP refers to an assessment criterion descriptor; the term refers to an area of communication in the PYP. With respect to the framework for language acquisition, we have illustrated that phases of communication, areas of communication, and communication continuums are highly similar in the PYP and MYP. By contrast, although the DP²⁶ reveals the spirit of these components in the language acquisition framework in the *MYP Guide*, the latter two programmes in the continuum do not share common terminology. Furthermore, the term “communicative processes” is not included in either the PYP or the DP. Any conclusions regarding how these processes align among programmes have to be made through use of the assessment criteria listed. In the case of the PYP, no assessment criteria are defined, thereby complicating such judgments.

²⁶ For example, DP assessment tasks involve the visual, oral, and written areas of communication.

PHASE V: DISCREPANCY ANALYSIS (CONCLUSION)

After examining findings from the first four phases and data from the three follow-up interviews, we conclude this report by grouping findings into areas of convergence, complementarity, and contradiction. Last, we discuss next steps that IB might consider when addressing findings from this research project.

Convergence

Among several themes that emerged in this study, three converged across data sources: the MYP *Guide's* avoidance of a linear progression, its encouragement of learners to analyze a variety of text types, and its emphasis on contextualized communication. Findings from both the literature review and cross-document analysis supported the MYP *Guide's* contemporary approach in recognizing that L2 learning does not follow an exact progression to which all students conform. Second, those same data sources showed that MYP learners benefit from being expected to analyze various text types, which engenders communication within and beyond their classrooms and the negotiation of meaning in communicative processes. Third, all the data sources demonstrated that the MYP *Guide* emphasizes meaningful, contextualized communication, rather than lower-level skills such as grammatical competency. Progression analysis findings described higher-level skills as an “overwhelming focus,” a recipe for “ideal language learning” in authentic settings. However, two discrepant views suggested that though the MYP *Guide* does focus on communication acts, rather than syntactic or lexical accuracy, such a focus might not be desirable among all stakeholders. Cross-document reviewers noted that the MYP *Guide* could increase its emphasis on reducing students’ language use errors. One of the discrepancy analysis teacher interviewees recognized that she might “be old school, but I think that I still need” to teach vocabulary, grammar, and syntax directly. As a result, the teacher reported that there is “not always time to be in an authentic context every single day.”

Convergence of Key Themes

- Language acquisition is nonlinear
- Students benefit from being asked to analyze various text types
- MYP *Guide* emphasizes meaningful, contextualized communication over low-level skills (e.g., grammar)

Importantly though, these perspectives reflect educators’ preferences, not their impressions of the MYP *Guide*. In our follow-up interview with the Head of MYP Development, he validated this concern, noting that language acquisition at the early stages of development might prompt some educators to emphasize memorization of characters (e.g., the kanji in Japanese) or conjugation of verb tenses, actions that might “register as inauthentic” or feel “very school-like” rather than the communicative process. He recognized that some days require more routinized practice than others, but stated “IB’s constructivist bent” tilts toward authentic approaches rather than the more traditional model of recitation followed by examination. However, he added that the MYP *Guide* was designed purposefully “to challenge that model, but is not meant to dismiss its strengths or say it is never allowed to do such a thing.”

Complementarity

Eight themes that emerged in this study were complementary across data sources:

- **Clear definitions:** Cross-document analysis findings included praise for the *MYP Guide* in defining phases of language acquisition, which one of the discrepancy analysis interviewees confirmed. However, all our data sources contained at least one criticism about a lack of explicit direction for educators, inconsistent use of terminology, failure to differentiate between strands, or reluctance to provide illustrative examples. The same discrepancy analysis interviewee who praised the definitions also lamented that “the only thing the IB shows us are the assessments and not how to get there.”
- **An integrated approach:** Findings from the literature review and cross-document analysis both show MYP’s authentic, holistic, conceptual, and inquiry-based approach to be a strength, though the literature review also unearthed two challenges for teachers who use the *MYP Guide*: Integrative approaches make it harder to provide comprehensible student feedback and teachers might struggle to determine how continuums, processes, and skills interact.
- **Emphasis on intercultural competence:** Almost all data sources identified the emphasis on intercultural competence as a clear strength for MYP, including the appreciation for one’s own culture and that of others. Cross-document analysis showed a particular strength in embedding the IB philosophy and in being culturally inclusive. However, the same data source questioned the *MYP Guide*’s lack of focus on “critical cultural awareness.” Relatedly, we witnessed a debate across data sources in which within-document analysis identified responses to assessment tasks in students’ mother tongues to be problematic. By contrast, cross-document analysis found the *MYP Guide* to offer little explicit consideration of native students’ language needs. Meanwhile, one of the discrepancy analysis interviewees urged the separation of L1 and L2 learning because the developmental processes are quite different.
- **Scope:** Cross-document analysis responses showed overlapping and complex views of the scope the *MYP Guide* should maintain, specifically in comparison to ACTFL. One reviewer found ACTFL to be more flexible. The other reviewer found ACTFL to be more concise and less overwhelming than the *MYP Guide*, which she realized carried a wider scope, and wondered how one might make it more concise without sacrificing its beneficial complexity. The Head of MYP Development characterized the *MYP Guide* as “a framework, and a guide toward, not a full teaching specification, nor a complete textbook, nor a set of teaching materials that’s a one-stop shop, open up this and start your language course today.”
- **Effectiveness of the progressions:** There were shades of difference between how the literature review and within-document analysis characterized the progressions. In the literature, we found the MYP’s proficiency progression across phases to cohere with current scholarship, but no empirical studies support this model. Nor was there a consensus approach for how to operationalize such a progression. Not surprisingly,

Complementary Themes

- *MYP Guide*’s definitions are generally clear
- Series of overall strengths that could use minor tweaks: integrated approach, focus on intercultural competence, progressions, international alignment, balancing theory and practice, links to PYP and DP

within-document analysis data ranged from describing the progression as “effective” to “generally effective” but not “smooth” to having wide gaps between some phases.

- **Alignment with CEFR and ACTFL:** The literature review revealed indirect alignment with the two internationally ubiquitous frameworks, but the cross-document analysis showed the *MYP Guide* and the CEFR to be particularly strong in similar areas, especially interpersonal communication. Data indicated less clear alignment between the *MYP Guide* and ACTFL.
- **Balancing theory and practice:** The literature review revealed a combination of solid theoretical underpinnings and practical challenges, so it seems fitting that the within-document analysis revealed issues of appropriateness and logical progression.
- **Linking the MYP to the PYP and DP:** Generally, the data sources agreed that the MYP links well to the programmes that precede and succeed it in the IB continuum. Findings from the progression analysis illustrate these connections well; within-document reviews and discrepancy analysis interviews noted strong connections with the DP in particular. However, the progression analysis highlighted differences in terminology, frameworks, and assessment approaches across the three programmes. Phase V interviewees expressed varying opinions about the extent to which those discrepancies might be problematic. For example, one teacher at a continuum school noted the challenge for students and parents to adjust when switching from MYP to DP assessment expectations. The same teacher identified a struggle for teachers who only teach in one programme to understand how the others operate. Furthermore, the teacher noted the challenge of knowing “where PYP students are coming from because they don’t have the same kind of assessment.” By contrast, the other teacher interviewee saw no such issues and had her own conflicting opinions about whether terminology and assessment should be uniform across programmes. She said,

In some ways, I think it will be easier perhaps for some teachers that don’t have the know-how, but on the other hand, in the DP you are usually working with grown-up kids and having a different approach, using different terminology, perhaps it’s not so necessary. . . . I understand that for concept-based learning it would help, the Approaches to Learning skills also, but I don’t think that everything must be repeated at the DP. It would be beneficial for some schools to use the same terminology. . . . Subtle changes in terminology do not affect learning or classroom approaches. It’s a matter to be considered, especially for those teachers who are new or don’t have experience in the programme.” (Interview 2).

Contradiction

Four themes that emerged in this study revealed contradictions between data sources: (a) emphasis on both formative and summative assessment, (b) using assessments to place students, (c) describing the proficiencies and communicative processes in the *MYP Guide*, and (d) incorporating digital and visual literacy. We discuss each of these themes with the overall caveat that although much of our data comes from reviewers with expertise in designing or implementing IB language acquisition curriculum, interpretations can vary by reviewers’ levels of experience. We strove to involve experts with experienced knowledge, not to capture expert-to-novice differences that might accompany various stakeholders’ readings of a document such as the *MYP Guide*.

First, the literature review highlighted the importance of incorporating both formative and summative assessment, but cross-document analysis findings showed the CEFR to be more adept in that area than the *MYP Guide*. This discrepancy likely occurred due to differences between an analysis grounded in a potentially aspirational literature base and the perspectives of IB educators and staff who see the day-to-day applications of those aspirations.

Second, the within-document analysis showed the assessment criteria, continuums, and other *MYP Guide* components to align with one another and facilitate judgments about student placement and instructional planning. Instead, the literature review pointed to the problematic issue of learners engaging in *either* language input or output on assessments, which could complicate feedback loops. Furthermore, within-document reviewers also took issue with the word counts in assessment tasks and their levels of challenge. Both within- and cross-document analysis findings raised the issue of teachers needing additional support to help differentiate students by ability, the inappropriateness of assessing students at the same range across skills, or how to promote students between phases with evaluative certainty. Last, one within-document reviewer questioned whether the MYP phases adequately prepared students to progress to the DP's Language A course. One of the two discrepancy analysis interviewees agreed that students who come from MYP as Phase 5 or 6 learners frequently must drop from the Language A or Language B HL courses to Language B SL. The differences noted here might reflect gaps between an overall positive trend that reviewers identified quantitatively alongside particular issues that can only be unpacked using qualitative methods.

Areas of Contradiction

- Interpretations of *MYP Guide's* emphasis on formative and summative assessment
- *MYP Guide's* usefulness for helping schools place students
- Descriptions of proficiencies and communicative processes
- Perspectives on incorporating digital and visual literacy

Third, both the literature review and the cross-document analysis describe the *MYP Guide's* presentation of the language proficiency and communicative competence as a strength. By contrast, other data from the cross-document analysis suggest gaps in strategic competency and listening comprehension, the latter of which a discrepancy analysis interviewee confirmed. This discrepancy seems like a function of examining the general case and the specific. On the whole, the presentation of language proficiency and communicative competence seem to be a strength, but strategic competency and listening comprehension seem to offer particular opportunities to improve the *MYP Guide*.

Fourth, though the literature review justified the inclusion of digital and visual communication in the *MYP Guide*, within- and cross-document analyses showed an "overemphasis" on visual skills when compared to speaking or written skills and weaknesses in the *MYP Guide* to foster students' abilities in "digitally mediated environments," "face-to-face interactions," or "synchronous interactive writing" such as text messaging. Given that the cross-document analysis noted gaps in these areas for the CEFR, ACTFL, and HKEB, it seems that all these language documents can improve in modernizing and expanding their approaches to defining literacy so they incorporate digital and visual elements.

Next Steps

In this final section, we present four questions that IB might consider when considering approaches to revising the *MYP Guide*. Text boxes reflect the key tensions that underscore those questions. To aid reflection on these questions, we present insights from the Head of MYP Development. Across all four questions, the Head of MYP Development seeks balance between what is ideal and what is feasible when considering steps forward from the project:

We can imagine and create a framework for a course that is innovative, or at least current. . . . We can build that and make it very sophisticated and contemporary, but that has to stand in balance with its teachability and assessability, which sometimes stand in pretty strict contradiction to what's real.

1. How can the *MYP Guide* balance exemplification and full detail with simplification and clarity?

This study found ample evidence to recommend corrections of imprecise uses of terminology and standardization that would address inconsistent application of those terms, both of which seem to reduce clarity in portions of the *MYP Guide*. Relatedly, reviewers sought

- more explicit definitions of terms
- examples that illustrate expectations (e.g., how educators might conceptualize a proficiency-based classroom)
- a combination of appropriateness and logical consistency
- sharper differentiation between strands

Key Tensions

- Which terms to define explicitly?
- How much exemplification to provide?
- How to sharpen differences between strands without overprescribing?

The Head of MYP Development understands the “common desire for schools to want to know exactly what they should do,” but distinguished exemplification of what a criterion “might look or sound like” with the type of highly contextualized information necessary for making specific pedagogical decisions. He noted that the type of comprehensive product or degree of standardization required for a fully exemplified guide would be “neither desirable, nor possible,” especially given that the *MYP Language Acquisition Guide* is “already 75% longer than any other guide” due to the multicourse nature of L2 instruction. Fully explicating every possibility to compare students would require an extensive bank of examples that cater to as many as 20 languages. Multiplying that number of languages by six phases and then by the other components of the *MYP Guide* would take an effort that “vastly exceeds IB’s [current] capacity.”

That being said, he noted that the IB continues to work toward more clearly specifying, clarifying, and removing redundancy and overlap in its guides (e.g., phases, progressions, processes, continuums) that are “all good things” but might be overly abundant and would yield additional complexity that might be burdensome for many teachers, especially those of entry-level language learners. He noted the importance of consistent use of terms.

Key Tension

- How to balance needs for consistency and differentiation?

2. To what extent should language, assessment criteria, and other facets align across the IB continuum?

To align or not to align the PYP, MYP, and DP on these bases may not be the question. Instead, how IB can achieve requisite alignment in key situations might be a more important consideration. The progression analysis and some reviewers suggested a need for shared language across programmes, an opinion that was not uniform among IB staff or MYP educators.

Clearly, this area holds significant room for various opinions. Contemporary language acquisition research does not provide a strong basis to validate one programme's approach to executing the foundational principles of language acquisition and assessment over the other IB programmes. However, considerable differences between PYP, MYP, and DP may complicate learner and practitioner understanding of the expectations inherent to each programme as they participate across the continuum. Still, the extent to which the programmes should align their terminology, assessment criteria, and communicative processes remains an open question.

The Head of MYP Development identified a tension between (a) reasons for IB's programmes maintaining their distinctive developmental focus and (b) a shared commitment for programmes to continue working and growing together as a seamless but developmentally differentiated continuum. In terms of shared language, he said that such a commitment is already in place, though the process is time-consuming and complex. In terms of sharing assessment criteria, he cited developmental considerations between programmes. For example, taking up a language for first time at age 3 when starting PYP differs from starting MYP at age 11 or DP at age 16 or learning a language as an adult, so "the same family of skills . . . might look different" developmentally. For this reason, variation in assessment tasks was "not a negotiable point" because as students mature along with their capabilities "they can do more and different things." He noted the limited definition of assessment in the PYP, a mixed level in MYP, and highly specified university-entrance specification in the DP. Of crucial importance, though, should be IB's attention to transitions between programmes, the range of goals that those programmes allow students to pursue, and ensuring reasonable pathways to those goals. Last, he highlighted the Career-Related Programme—which was not a target of this study—as having its own language-acquisition requirements that differ from those of the DP, an example of how variation for reasons other than developmental trajectory are also important considerations.

3. What assessment changes can the IB undertake without creating additional challenges for practitioners and schools?

During this study, several instances occurred in which reviewers or analysts identified opportunities to enhance MYP's approaches to assessment. For one, several reviewers asked the MYP to reexamine word-length ranges on assessments. The Head of MYP Development described such a fix as being fairly easy to do, requiring some "highly technical" considerations, but a change that could benefit from the participatory nature of IB's curriculum review process where teacher input is intentionally

Key Tensions

- Reconciling word lengths on assessments?
- Integrating mother tongue?
- Merging assessments to incorporate interpretation and production?
- Technological solutions that maintain authentic assessment?

integrated. Another reviewer asked about reconsidering the appropriateness of using mother tongue and/or language of instruction during summative assessments.

Other assessment changes might require greater lifts. For example, literature review findings pointed toward merging assessments, allowing language learners to engage in both input and output on a single assessment, thus avoiding complications in teacher-to-student feedback loops. Meanwhile, other data raised a tension between newer assessment modes, such as those in digitally mediated environments or synchronous interactive writing, and older modes such as attending to errors in vocabulary, grammar, and syntax. The Head of MYP Development considered both of these issues to be solvable given expected improvements in the logistical and technical possibilities of using computers effectively to capture students' productive speech. He was confident that IB would move in that direction, as long as standardizing the assessment were possible.

However, he cautioned about the gap between what is innovative from a pedagogical standpoint and what might be feasible from an assessment standpoint. He conceived of a scenario in which a language acquisition course esteemed the most authentic communication environment, but could not develop sufficient assessments to match. In that instance, one might travel to Paris and speak to locals in a restaurant or café: "You'd be able to conduct yourself well in a culturally appropriate manner and understand all the transactions going on there, but you might still get a bad grade in French if the assessment was designed for a different set of objectives." Similarly, he placed a high value on a Skype conversation with a student in another country where the target language was spoken. A teacher might be able to determine generally how the student performed, but "how you are doing generally isn't always assessable reliably in a summative sense and it is quite complicated to arrange that teaching and learning to assess it beyond just saying you had a great conversation with your pen pal."

4. What else can the MYP Guide borrow from the other international language frameworks?

Though the MYP Guide has definitely leaned upon international language frameworks, most notably the CEFR and ACTFL, reviewers identified additional areas where the MYP could still benefit. According to findings from the cross-document analysis, MYP could look to the CEFR for help in the areas of strategic competency, summative and formative assessment, communication in face-to-face interactions and digitally mediated environments, and helping teachers attend to students' errors. Reviewers also indicated that the ACTFL and HKEB both provided clearer descriptions and richer examples, which might or might not be advantageous in thinking about revision for the MYP Guide. Last, both reviewers and interviewees suggested that a crosscheck

between the MYP Guide and sources such as the HKEB and the IB's documents on Approaches to Teaching and Learning would increase MYP focus on some dispositional aspects that are at the core of the IB mission.

In total, these findings reflect contributions of MYP internal staff, educators in IB schools from five countries, and external researchers. Had we conducted only a literature review, we would have concluded that increasing specificity and exemplification should inform IB's next steps.

Key Tensions

Improving focus on

- Strategic competency (see CEFR)
- Summative and formative assessment (see CEFR)
- Attending to student errors (see CEFR)
- Use of clear descriptions and rich examples (see ACTFL and HKEB)

But this study supports the strength of a mixed-methods evaluative approach. Examining literature review findings alongside results from within- and cross-document analyses, it seems clear that IB needs to attend to specificity and exemplification in some areas, but not universally across its language acquisition guide. After accounting for findings from the progression and discrepancy analyses, we recommend that IB begin conversations about the scope and purpose of change before making any decisions about what in the *MYP Guide* requires revision, how to approach that revision, and to what extent revision is advisable.

REFERENCES

References include only works cited in the final report. The annotated bibliography (Appendix A) includes a list of references that comprehensively includes the works included in that document.

- Alderson, J. C. (2007). The CEFR and the need for more research. *The Modern Language Journal*, 91(4), 659–662.
- American Council on the Teaching of Foreign Languages (ACTFL). (2011). *Standards for foreign language learning in the 21st century*. Alexandria, VA: Author.
- Atkinson, D. (2002). Towards a sociocognitive approach to second language acquisition. *The Modern Language Journal*, 86(4), 525–545.
- Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Clevedon, UK: Multilingual Matters.
- Byram, M., Nichols, A., & Stevens, D. (2001). Introduction. In M. Byram, A. Nichols, & D. Stevens (Eds.), *Developing Intercultural Competence in Practice* (1-8). New York, NY: Multilingual Matters.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47.
- Council of Europe. (2002). *Common European Framework of Reference for Languages: Learning, teaching, assessment. Case studies*. Strasbourg, France: Author.
- Curriculum Development Council and the Hong Kong Examinations and Assessment Authority. (2015). *English Language Education Key Learning Area English Language Curriculum and Assessment Guide (Secondary 4–6)*. Retrieved from <http://www.edb.gov.hk/en/curriculum-development/kla/eng-edu/curriculum-documents.html>
- Devitt, S. (1997). Interacting with authentic texts: A multilayered process. *The Modern Language Journal*, 81(4), 457–455.
- Gass, S., & Selinker, L. (2008). Input, interaction, and output. In *Second language acquisition: An introductory course* (3rd ed., pp. 304–367). New York, NY: Routledge.
- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation design. *Educational Evaluation and Policy Analysis*, 11(3), 255–274.
- Harklau, L. (2002). The role of writing in classroom second language acquisition. *Journal of Second Language Writing*, 11, 329–350.
- Herrera-Viedma, E., Cabrerizo, F. J., Kacprzyk, J., & Pedrycz, W. (2014). A review of soft consensus models in a fuzzy environment. *Information Fusion*, 17, 4–13.
- Hulstijn, J. H. (2007). The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. *The Modern Language Journal*, 91(4), 663–667.
- Hulstijn, J., Schoonen, R., Jong, N., Steinel, M., & Florin, A. (2011). Linguistic competences of learners of Dutch as a second language at the B1 and B2 levels of speaking proficiency of the Common European Framework of Reference for Languages. *Language Testing*, 29(2), 203–221.
- Hyland, K. (2007). Genre pedagogy: Language, literacy, and L2 writing instruction. *Journal of Second Language Writing*, 16, 148–164.

- International Baccalaureate Organization. (2014). *Middle Years Programme: Language acquisition guide*. Cardiff, Wales.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174. doi: 10.2307/2529310.
- Liskin-Gasparro, J. E. (2003). The ACTFL Proficiency Guidelines and the Oral Proficiency Interview (OPI): A brief history and analysis of their survival. *Foreign Language Annals*, 36(4), 483–490.
- Little, D. (2007). The Common European Framework of Reference for Languages: Perspectives on the making of supranational language education policy. *The Modern Language Journal*, 91(4), 645–655.
- Lotherington, H., & Jenson, J. (2011). Teaching multimodal and digital literacy in L2 settings: New literacies, new basics, new pedagogies. *Annual Review of Applied Linguistics*, 31, 226–246.
- Mantero, M. (2002). Evaluating classroom communication: In support of emergent and authentic frameworks in second language assessment. *Practical Assessment, Research, & Evaluation* 8(8), 1–7.
- Merriam, S. B. (1998). *Qualitative research and case study applications in education (2nd ed.)*. San Francisco, CA: Jossey-Bass.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–756.
- Nevo, B. (1985). Face validity revisited. *Journal of Educational Measurement*, 22(4), 287–293.
- North, B. (2007). The CEFR illustrative descriptor scales. *The Modern Language Journal*, 91(4), 656–659.
- Porter, A. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, 31(7), 3–14.
- Savignon, S. (1983). *Communicative competence: Theory and classroom practice*. Reading, MA: Addison Wesley.
- Siegel, M. (1996). The role of learner subjectivity in second language sociolinguistic competency: Western women learning Japanese. *Applied Linguistics*, 37(6), 356–382.
- Spada, N. (2005). Conditions and challenges in developing school-based SLA research programs. *The Modern Language Journal*, 89(3), 328–338.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. *Input in Second Language Acquisition*, 15, 165–179.
- Westhoff, G. (2007). Challenges and opportunities of the CEFR for reimagining foreign language pedagogy. *The Modern Language Journal*, 91(4), 676–679.
- Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *The Phi Delta Kappan*, 70(9), 703–713.
- Wiggins, G. (2011). Moving to modern assessments. *The Phi Delta Kappan*, 92(7), 63.
- Yuksel, D., & Inan, B. (2014). The effects of communication mode on negotiation of meaning and its noticing. *ReCALL*, 26(3), 333–354.

APPENDIX A: ANNOTATED BIBLIOGRAPHY

This annotated bibliography has three sections:

1. Resources beneficial to evaluating the phases of language acquisition exemplified in the global proficiencies table, language continuums, and assessment criteria.
2. Resources beneficial to evaluating the communicative processes set forth in the assessment criteria and to consideration of how to blend the four language continuums.
3. Resources beneficial to considering the assessment design of the *MYP Language Acquisition Guide*.²⁷

I. Phases of Language Acquisition

Alderson, J. C. (2007). The CEFR and the need for more research. *The Modern Language Journal*, 91(4), 659–662.

Highlights:

1. *Second Language Acquisition and Testing in Europe (SLATE) is working toward developing a learner corpus to inform the validity of Common European Framework of Reference for Languages (CEFR).*
2. *Criticizes textbooks manufacturers for aligning their products to the CEFR due to its lack of empirical support.*

Alderson calls for more research regarding the CEFR and its use. The CEFR makes no reference to specific languages, assuming that any communicative task demands a comparable level of proficiency between languages. Although some reference-level descriptions are being created, Alderson notes that the methods for that work are “unclear or suspect” (p. 660). He contends that fundamental research is needed into how proficiency in the main European languages develops over time. He suggests a carefully built European learner corpus, such as an attempt by SLATE. He criticizes how examination providers, textbook publishers, and curriculum developers often connect their products to the CEFR, despite the presence of gaps and flaws. He argues for the creation of a body with authority to examine such claims critically. The article advocates a revised CEFR that accounts for current criticisms and limitations.

Byrnes, H. (2007). Introduction. *The Modern Language Journal*, 91(4), 641–645.

Highlights:

1. *Plurilingualism contrasts with multilingualism in that communication is valued from all languages (plurilingualism) rather than just major languages (multilingualism).*
2. *The article serves as an introduction to a series of articles designed to discuss the CEFR. These articles are annotated in this bibliography.*

In this article Byrnes introduces topics of debate around CEFR, including scholars’ perspectives. He writes that as globalization continues to spread, governments have a responsibility to address complex educational issues and argues that the United States (US) has not done this (though national security concerns have spurred some interest). By contrast, the European Union (EU)

²⁷ Some resources listed under Section II: Communicative Processes and Language Continuums are also relevant to Section III: Assessment because researchers sometimes consider communicative processes and the evaluation of those processes simultaneously.

has advocated strongly for and promoted language teaching and learning, proclaiming a policy of plurilingualism. As later discussed by Little, plurilingualism differs from multilingualism by highlighting the way that plurilinguals can communicate with any of the languages available to them, not just the major ones. Byrnes writes that EU policies, unlike those in the US, react to and position themselves within larger sociocultural trends. Such recommendations reflect the realities of the governing body, rather than the ideal. The article provides (a) a multiperspective overview of how CEFR frames language education in Europe, (b) future directions for focus, and (c) reflections on process and product dynamics in a policy-setting context, all to address the inappropriate application of CEFR to other contexts and to generate ideas for how the challenges might be addressed in new environments with similar multilingual realities. Byrnes also says that there have been many successes with CEFR, though the enlargement of the EU places CEFR at a crossroads. She then introduces the contributors, who are annotated within this bibliography.

Clifford, R., & Cox, T. L. (2013). Empirical validation of reading proficiency guidelines. *Foreign Language Annals*, 46(1), 45–61.

Highlights:

1. *There are differences in difficulty between the American Council of Teaching Foreign Languages (ACTFL) levels for reading proficiency, and those levels are hierarchical.*
2. *Blended task and text combinations (when texts and reading tasks do not align in their appropriate application to a given proficiency level) yield inaccurate assessments of proficiency level.*

In this study, Clifford and Cox explored the validity of the ACTFL reading proficiency guidelines, testing whether the receptive skill levels could be described in a hierarchical order of difficulty. The authors tested in English, Spanish, and Chinese, and considered reading assessment items clustered by level and order. For all three languages, they found empirical evidence that intended proficiency level of test items had a strong effect on item difficulty. This study supports the notion of hierarchical difficulty differences between levels, supporting the ACTFL reading proficiency guidelines. Implications for teachers include a need to look at each level as requiring a unique set of assessment criteria and ensuring alignment between those criteria and the reading texts. For example, if a learner can understand the main idea of an advanced text (an intermediate-level skill), the learner should not be attributed with advanced-level abilities. Teachers should also scaffold based on difficulty.

Fulcher, G. (1996). Invalidating validity claims for the ACTFL oral rating scale. *System*, 24(2), 163–172.

Highlights:

1. *Dandonoli and Henning (1990) and Henning (1992) do not show evidence of validity for ACTFL oral rating scales in the way that they claim.*
2. *ACTFL scales lack empirical evidence and should have had validity verified during development, rather than after.*

Fulcher examines papers from Dandonoli and Henning (1990) and Henning (1992), arguing a lack of sufficient evidence in those papers to claim validity in ACTFL's oral rating scale. Fulcher criticizes the lack of empirical evidence in developing the scale and questions the methods used and conclusions gleaned in the prior studies. Fulcher states that validity should be assessed during assessment development, rather than after.

Granfeldt, J., & Agren, M. (2014). SLA developmental stages and teachers' assessment of written French: Exploring Direkt Profil as a diagnostic assessment tool. *Language Testing*, 31(3), 285–305.

Highlights:

1. *Direkt Profil is a computer program that analyzes the developmental stage of second language (L2) French learners.*
2. *When Direkt Profil assessment is compared with trained language teachers' assessments for lower level and higher-level learners, the teachers and Direkt Profil had a high correlation; however, there was more variation when it came to intermediate level learners.*

The authors briefly discuss a theory of L2 French development from Bartning and Schlyter (2004), in which six stages of development take on grammatical profiles. Direkt Profil users submit a sample of interlanguage production, such as a written text or a transcription of oral production, and then Direkt Profil maps out the development level as defined by Bartning and Schlyter (2004). This study selected 50 student-produced written texts for evaluation by Direkt Profil. Then, experienced teachers assessed the texts. Teachers did not receive any special instructions on assessment procedures. After assessing the texts, teachers were asked to state their degree of certainty for their evaluations. No teacher had median assessment scores in which form (syntactic and lexical accuracy) was marked higher than content (overall message), indicating that form usually scores lower than content in written assessments. When considering proficiency, this suggests learners may be able to begin to deal with complex topics outside of their proficiency levels before acquiring the lexical and syntactic skills to do so without error. It is important to note that there was also greater interrater variability in the assessment of content, which could indicate subjectivity in teachers' ratings.

In the study, there was a high correlation between the mean values of the teacher assessment scores and the developmental stage as defined. The authors reported a strong correlation between Direkt Profil's grammatical development stage with teacher assessment scores than with a learner's year of language acquisition study, indicating the within-classroom range of proficiency levels most teachers serve. The authors conclude that although nothing can replace a teacher's evaluation, programs such as Direkt Profil could "ease some of the heavy burden of diagnostic assessment of form that weighs on language teachers in the school context" (p. 303). This diagnostic assessment is particularly important in educational contexts in which a learner's proven proficiency level determines his or her classroom placement.

Hulstijn, J. H. (2007). The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. *The Modern Language Journal*, 91(4), 663–667.

Highlight:

1. *Corpus research is needed to empirically evaluate the CEFR.*

Hulstijn begins by praising strengths of the CEFR and then focuses on how improvements need to be approached. She discusses how the CEFR looks at quantity and quality in the proficiency descriptors. However, there has not been enough empirical research to say whether or not "a learner at a B2 level of overall production must also have attained the B2 level on all the linguistic competences²⁸ and scales, or whether it is possible for a learner to be situated at different levels

²⁸ In the annotated bibliography (Appendix A) and full-length literature review (Appendix B), we default to each author's use of competence or competency.

on different scales” (p. 664). She identifies a need to improve the CEFR theoretically “using real L2 learners rather than teachers for rank ordering descriptors,” because no longitudinal studies have shown that all L2 learners first attain the functional level of A1, then A2, B1, B2, C1, and C2.²⁹ Also, no research indicates that learners beyond an A1 level are able to perform all tasks from the lower level (p. 666).

Hulstijn, J., Schoonen, R., Jong, N., Steinel, M., & Florin, A. (2011). Linguistic competences of learners of Dutch as a second language at the B1 and B2 levels of speaking proficiency of the Common European Framework of Reference for Languages. *Language Testing*, 29(2), 203–221.

Highlights:

1. *No evidence suggests that a learner performing at a given level on the CEFR scale actually possesses the linguistic competences for that scale: “The CEFR levels, in their present form, are neither based on empirical evidence taken from performance, nor on any theory in the fields of linguistics or verbal communication” (p. 204).*
2. *Threshold (B1) learners demonstrated vocabulary breadth way above that which would be predicted of their proficiency levels, making the need for more empirical evidence to support what level of vocabulary knowledge learners need at a given proficiency level.*

The authors examine how L2 learners’ skills equate to their CEFR score, focusing on the differences between B1 and B2 learners specifically. They gave 181 adult learners of Dutch eight speaking tasks to find out how a learner’s knowledge of vocabulary and grammar affect their scores. They asked if Dutch L2 learners differ only in their knowledge of low-frequency words or in their knowledge of words in all frequency classes to the same degree (p. 205). They found somewhat larger B1–B2 learner gaps for medium- and low-frequency words compared to high-frequency words, making vocabulary seem to be the most important linguistic component at lower levels. This finding indicates a need for more research to support understanding of the target vocabulary level of knowledge needed for each proficiency level. A limitation of this study was the use of simulated conversations.

Krumm, H. J. (2007). Profiles instead of levels: The CEFR and its (ab)uses in the context of migration. *The Modern Language Journal*, 91(4), 667–669.

Highlight:

1. *The application of the CEFR to allow for citizenship in some countries is inappropriate both for the high-stress examination session that it engenders and its inadequate consideration of the language profile of people groups who were not considered in the framework’s development, particularly migrants.*

²⁹ CEFR levels are as follows: A1: Breakthrough, or a basic ability to communicate and exchange information in a simple way; A2: Waystage, or an ability to deal with simple, straightforward information and begin to express oneself in familiar contexts; B1: Threshold, or the ability to express oneself in a limited way in familiar situations and to deal in a general way with nonroutine information; B2: Vantage, or the capacity to achieve most goals and express oneself on a range of topics; C1: Effective operational proficiency, or the ability to communicate with the emphasis on how well it is done, in terms of appropriacy, sensitivity and the capacity to deal with unfamiliar topics; C2: Mastery, or capacity to deal with material which is academic or cognitively demanding, and to use language to good effect at a level of performance which may in certain respects be more advanced than that of an average native speaker.

Krumm shows the application of CEFR to groups of people who were not emphasized during the framework's development, specifically migrants. Some European countries require migrants to master a certain level of language before becoming a citizen. Krumm illustrates that such high stress is doubtful to provide a good basis for successful language learning. Furthermore, most CEFR descriptors do not relate to migrants' sociocultural contexts. Krum notes some underlying assumptions of descriptors ignore or clash with migrant experiences. Also, not all migrants will need the same level of proficiency in all areas to be successful, depending on vocation and/or situation. As Krumm explains, "Thus, a migrant with a profile of B1 speaking, B2 listening, A2 reading and A1 Writing might be very successful in his or her work and well integrated into the new society" (p. 668). The CEFR acknowledges that native speakers show uneven competencies.

Leung, C., & Lewkowicz, J. (2013). Language communication and communicative competence: A view from contemporary classrooms. *Language and Education*, 27(5), 398–414.

Highlights:

1. *The current definition of communicative competence as used in existing frameworks is limited in that it doesn't include all the functions that learners use.*
2. *Language that shows certain types of (currently unattended to) participation, engagement, and social interaction needs a place in curriculum and assessment.*

In this paper, Leung and Lewkowicz showed the results of analyses of three extracts of classroom discourse with native and nonnative English speakers. The first extract showed a speaker using a voluntary offer of support; the second showed a student deviating from the flow of the conversation with a higher-power native speaker to clarify information after the native speaker had signaled that the conversation had ended; the third used examples to present a complex point of view to contribute to a discussion without a main argument. Through the analysis of these speech functions used by nonnative speakers, the study shows that the CEFR does not provide descriptors of all the possible functions that learners may use, limiting the content validity of the CEFR. The authors conclude that curriculum design and assessment needs an expanded notion of communicative competence that allows for social interaction and includes both participation and engagement.

Lim, G. S., Geranpayeh, A., Khalifa, H., & Buckendahl, C. W. (2013). Standard setting to an International Reference Framework: Implications for theory and practice. *International Journal of Testing*, 13, 32–49.

Highlights:

1. *Due to the nature, goals, and context of the CEFR, standard setting poses challenges.*
2. *The results from a standard-setting study and criterion validation study agree and provide similar CEFR level descriptions.*

The authors examine the relation between standard setting and its application to international reference frameworks, such as the CEFR with its wide variability and diverse contexts. The typical distinction in standard setting between content and performance standards doesn't apply with CEFR. Moreover, different test providers that link to the CEFR obtain divergent outcomes. The authors present two studies, a standard-setting study and an external validation study. Their findings across studies support the idea of convergent outcomes in standard setting and show that two different groups working with two different exams can agree on multiple CEFR levels that are comparable.

Liskin-Gasparro, J. E. (2003). The ACTFL Proficiency Guidelines and the Oral Proficiency Interview (OPI): A brief history and analysis of their survival. *Foreign Language Annals*, 36(4), 483–490.

Highlights:

1. The ACTFL guidelines led to introspection and innovation in language teaching, which helped ingrain pedagogy; the guidelines and OPI are now integrated into teaching, assessment, programming, and research.
2. The ACTFL guidelines and OPI have been heavily criticized in terms of empirically based validity, content choice, reliability, language facet focus, among others.
3. The much-criticized “non-technical and atheoretical stance” on proficiency that the ACTFL guidelines support has proliferated much of language education.

Liskin-Gasparro presents a 50-year history of the ACTFL guidelines from their development in the State Department, criticisms of the standards, and reasons for their ongoing relevance. Insights from teachers who examined their own practices in comparison to what students were required to do on the OPI ignited innovation in the field and greater focus on communication. Many criticisms of the guidelines also exist, including (a) the lack of empirical tests of validity, (b) the circularity of the guidelines, and (c) the use of the native speaker as a basis of measurement. Liskin-Gasparro criticized the OPI’s interrater reliability and the nature of the language that the OPI did (or did not) elicit, as well as its emphasis on grammar and vocabulary over other facets of language such as the consequences of power dynamics in the interview. The guidelines and OPI are viewed as major catalysts for change. Combined with the growth of ACTFL, this view facilitated the institutionalization of the guidelines, including their relation with large-scale tests and access provided to speech samples for OPI research.

Little, D. (2007). The Common European Framework of Reference for Languages: Perspectives on the making of supranational language education policy. *The Modern Language Journal*, 91(4), 645–655.

Highlights:

1. The CEFR is an optional, international, plurilingual-oriented language learning framework that has been implemented with mixed results regarding gathering of evidence to prove language proficiency.
2. The CEFR has little empirical backing, but it is beneficial in that its proficiency descriptors are clear and independent.

Little reviews the CEFR for (a) its analysis of what is involved in language use and learning; (b) its six levels of ability; (c) its lack of language specification; (d) its use of general competencies (knowledge, skills, etc.) and communicative language competences (linguistic, pragmatic, etc.); (e) its four categories of language activity (i.e., reception, production, interaction, and mediation) and four categories of language use (i.e., personal, public, occupational, educational); and (f) its three parameters (i.e., situational context, text type/theme, and conditions/constraints). Little concludes that the CEFR is action-oriented, meaning that language users are seen as social agents with tasks to accomplish; however, the scales are not solely communicative (including things such as planning, repair, etc.). Little recommends that users choose the scales most relevant to them. Little describes CEFR’s proficiency descriptors as always positive, definite, clear, and independent. Still, a variety of challenges are associated with the CEFR, including the lack of

empirical evaluation of the proficiency scales, its focus on the L2 instead of the more plurilingual (L1-centered) approach advocated by the Council of Europe, and the CEFR's inattention to differences among languages.

Martin, C. L. (2010). Assessing the oral proficiency of adult learners, "heritage" and "native" speakers using the "ILR Descriptions" and "ACTFL Proficiency Guidelines": Considering the challenges. *Russian Language Journal*, 60, 167–181.

Highlights:

1. *Adult foreign language learners, heritage speakers, and native speakers are assessed using the ACTFL and ILR (Interagency Language Roundtable) guidelines.*
2. *Although learners may have different skills, profiles, and differences based on heritage/native/foreign learner status, the guidelines for both ACTFL and ILR are based on function.*
3. *There should not be different assessments for native speakers than nonnative speakers.*
4. *A speaker's profile should not matter to end users.*

In this article, Martin answers questions about implementing ACTFL and ILR proficiency measures with different types of learners from varying language backgrounds (i.e., native, heritage, and other learner). Martin examines differences and similarities that these learners might encounter when rating language abilities in both contexts. Martin calls for discussion on the growing linguistically diverse group of language users in the United States. Thus, ensuring the appropriate evaluation of learners and the applicability of currently used scales is necessary.

North, B. (2007). The CEFR illustrative descriptor scales. *The Modern Language Journal*, 91(4), 656–659.

Highlights:

1. *The CEFR attempts to describe the vast, complex process of language acquisition and falls short: no empirical evidence supports the progression that it outlines.*
2. *No evidence suggests that the CEFR has increased the reflective practice of practitioners, though practitioners are beginning to share interpretations of the learning expectations set forth by the CEFR.*

North addresses Little's (2007) criticisms of the CEFR, asserting that Little tries "to summarize so much in so few words he, perhaps, does not do full justice to the flexibility of the CEFR" (p. 656). North describes the historical background of how and why the CEFR and the European Language Portfolio (ELP) were created, listing CEFR's descriptors that lack an empirical base and descriptors that draw the most user complaints. North highlights the vast undertaking that the CEFR is attempting to encompass, suggesting an expectation that it may not reach all its goals. North concludes that of the CEFR's three aims—"establish a common metalanguage to talk about objectives and assessment; encourage practitioners to reflect on their current practice, particularly in relation to analyzing practical language learning needs, setting objectives, and tracking progress; and agree on common reference points based on the 1970's work on objectives"—it achieves its first aim and the third aim is on its way to achievement (p. 659). North shares Little's disappointment with the extent and quality of reflection in Aim 2.

Smith, M. S., & Truscott, J. (2005). Stages or continua in Second Language Acquisition: A MOGUL solution. *Applied Linguistics*, 26(2), 219–240.

Highlight:

1. *Based on a review of different theories of stages and continua in acquisition research, learners progress slowly and move through different developmental stages.*

The authors discuss different theories about stages of language acquisition, notably Corder (1977), Selinker (1972), Brown (1973), Dulay et al. (1982), Slima and Bellugi (1967), Roberston and Sorace (1999), and White (1982, 2003), and come to a synthesis of language acquisition stages and continua, mainly that second language grammar is “continuous and nondiscrete, with extended periods of optionality” (p. 235). They rely on the Modular On-Line Growth and Use of Language (MOGUL) framework to explain their synthesis, which features explicit and direct connection between “activation levels” (p. 230). MOGUL captures second language learners’ slow progression that proceeds through different developmental stages.

Spada, N. (2005). Conditions and challenges in developing school-based SLA research programs. *The Modern Language Journal*, 89(3), 328–338.

Highlights:

1. *The authors explore epistemological, technical, and ethical challenges related to L2 research in school settings.*
2. *A descriptive/observational phase is important to classroom research. Striving to connect research to practice trumps any pursuit of generalizability.*

Spada draws from her own experience and the work of van Lier (1988), Leeman (2003), Spada and Lightbown (1993), Harley (1989), Rosenshine and Furst (1973), and Clarke (1995) to examine considerations when conducting school-based second language acquisition (SLA) research. She highlights the importance of an observational phase before beginning any interventions in the classroom to gather crucial information about various classrooms, teachers, and students. She reviews goals for school-based research grounded in socialization and poststructuralism, including a focus on learners’ use of the language and how classroom experiences and interaction contribute to growth. She draws on Clarke’s (1995) argument that studies should be “particularizable” rather than “generalizable” to help teachers make connections between their lives and the research findings.

Surface, E. A., & Dierdorff, E. C. (2003). Reliability and the ACTFL Oral Proficiency Interview: Reporting indices of interrater consistency and agreement for 19 languages. *Foreign Language Annals*, 36(4), 507–519.

Highlights:

1. *The OPI (Oral Proficiency Interview) is reliable and can be used with confidence to assess proficiency.*
2. *The ACTFL rating scale functions well and shows reliability.*
3. *Language characteristics tested do not affect interrater reliability.*

In this study, the authors tested the OPI reliability for 19 languages based on 5,881 interviews. The study categorized languages according to testing density (i.e., frequency of assessment) and language difficulty, based on an English-speaking learner. There was a high level of interrater reliability and consistency. Neither testing density nor language difficulty affected interrater reliability (the interrater reliability scores ranged from 96% to 98% across languages in the study).

This result indicates that raters with high levels of training and practice are more consistent in their rating of student samples than their less-trained or less-practiced peers. The highest agreement for raters occurred on the extremities of performance (novice low and superior). Overall, the ACTFL scales functioned as intended; the OPI results proved to be a reliable assessment of language proficiency levels.

Westhoff, G. (2007). Challenges and opportunities of the CEFR for reimagining foreign language pedagogy. *The Modern Language Journal*, 91(4), 676–679.

Highlights:

1. *Language learning does not happen in the linear fashion reflected in most textbooks.*
2. *Chunking language may build proficiency faster than simple vocabulary acquisition.*

Westhoff shows the linear process common to most European textbooks, where grammatical items are presented in stepwise fashion. However, Westhoff writes that recent insights in SLA research, alongside intuition and observation, reveal that language is not learned linearly. For beginning learners, meaning must be processed before rule building and use. Westhoff estimates that this transition occurs between levels B1.2 and B2.1. He discusses lexical units, or “chunks” of language, which he states are often underestimated. Because of the capacity of working memory, chunks are a fast way for beginners to build proficiency. He concludes that aligning the CEFR framework to current teaching methodologies would be a small revolution.

II. Communicative Processes and Language Continuums

Atkinson, D. (2002). Towards a sociocognitive approach to second language acquisition. *The Modern Language Journal*, 86(4), 525–545.

Highlights:

1. *The cognitive perspective of SLA is limited and isolated; a more authentic alternative would account for the social nature and cognitive processes of language.*
2. *A sociocognitive approach to SLA is more connected to other disciplines and to society and action-taking in general.*

Atkinson argues for a sociocognitive approach to SLA and claims that the cognitive approach is detrimental by being overly isolated and void of the context needed to authenticate language tasks. He writes that interaction, which is fundamental to the acquisition of language, is absent from mainstream SLA. He argues that the application of hard science methods for studying SLA is misinformed. Atkinson writes that cognition is an isolated process of taking in information that cannot describe the entirety of language use and interaction. He argues against the Chomskyan view of language, saying instead that language is a rich resource for social interaction. A sociocultural view of language acquisition, then, would account for the importance of language in social interaction and in other disciplines and activities, in addition to the cognitive processes associated with SLA. One implication of this view is its value to the teaching side of SLA and how it works with learning.

Byram, M., Nichols, A., & Stevens, D. (2001). Introduction. In M. Byram, A. Nichols, & D. Stevens (Eds.), *Developing Intercultural Competence in Practice*. Clevedon, UK: Multilingual Matters.

Highlights:

1. This chapter introduces a book designed to help educators understand how to embed intercultural competence into their respective classrooms.
2. Intercultural competence comprises the skills of interpreting and relating, discovery and interaction, and critical cultural awareness.

This chapter embodies Byram (1997) and provides educators with a brief overview of intercultural competence. The authors discuss the attitudes and knowledge that all language learners possess and emphasize that learning that values the target cultures as equal to one's own must be encouraged. The authors provide an overview of the skill sets that language learners must develop in order to acquire intercultural competence. These skills are:

- *Interpreting and relating (savoir comprendre)*: ability to interpret a document or an event from another culture, to explain it and relate it to documents of one's own.
- *Skills of discovery and interaction (savoir apprendre/faire)*: ability to acquire new knowledge of a culture and cultural practices and the ability to operate knowledge, attitudes and skills under the constraints of real-time communication and interaction.
- *Critical cultural awareness (savoir s'engager)*: an ability to evaluate, critically and on the basis of explicit criteria, perspectives, practices and products in one's own and other cultures and countries. (Byram, Nichols, & Stevens, 2001, p. 1–8)

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47.

Highlights:

1. Canale and Swain provide their first iteration of a theoretical framework for communicative competence in L2 acquisition.
2. The framework provides a mechanism for evaluating communicative methodologies for language acquisition.

Canale and Swain examine and evaluate methodological and theoretical approaches (organized as either grammatical and/or communicative) to language acquisition. Through this evaluation, the superiority of the communicative approach is established (i.e., the conveyance of meaning is valued over grammatical accuracy), and a theoretical framework for communicative competence is posited. This framework includes grammatical competence (ability to communicate with accurate language), sociolinguistic competence (ability to use language appropriately in a given context), and situational competence (ability to use verbal and nonverbal communication strategies to address breakdowns in communication). In this original version of the framework, discourse competence (ability to combine form and meaning in a way that is cohesive, complex, and appropriate) was addressed within discussions of sociolinguistic competence but separated as its own competence by Canale (1983) in "From communicative competence to communicative language pedagogy."³⁰

³⁰ Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. C. Richards & R. W. Schmidt (Eds.), *Language and Communication* (pp. 2–27). New York, NY: Longman.

Chun, D. M. (2011). Developing intercultural communicative competence through online exchanges. *CALICO Journal*, 28(2), 392–419.

Highlights:

1. *This study examined two different forms of online interaction— asynchronous forum discussions and synchronous chats—in order to observe if the medium changed the style of language the participants used.*
2. *American students qualitatively asked more questions than the German students, and the asynchronous forum contained more linguistic variation and advanced language than the synchronous chats, which were more simple and informal.*

In this study, Chun examined how the choice of internet tools, specifically asynchronous forum discussions and synchronous chats, contribute to the style in which a learner produces language. Participating students responded to forum posts and had online chat conversations. U.S. students asked more questions than their German peers. In addition, the language employed in the synchronous chats was linguistically simpler and informal, whereas the language in the forum discussions was more linguistically diverse, challenging, and formal. The author confirmed findings from Thorne (2003): different internet tools are not neutral media and foster different syntactic and pragmatic styles. Similarly, Chun & Wade (2004) found in asynchronous forums that true interaction is lacking, unless learners are trained to interact that way.

Devitt, S. (1997). Interacting with authentic texts: A multilayered process. *The Modern Language Journal*, 81(4), 457–455.

Highlights:

1. *Little research connects reading to L2 development, but “the situation in which L2 readers are struggling with lower-level processes [word-level meaning] in order to create meaning from a text . . . is analogous to that of nonnative speakers involved in negotiation routines during spoken conversation” (p. 463).*
2. *Scaffolding reading processes for learners may be critical so that higher levels of text-based negotiation of meaning can exist in the L2, particularly for beginning learners.*

Devitt attempts to fill a void that exists within SLA literature regarding both reading processes and the ways that reading might drive language acquisition. Devitt posits that the fields of L2 input/interaction and L2 reading may be able to “fertilize” the other. He supports this claim by outlining that readers have textual interactions that are similar to interactions in conversations among L2 speakers, though the former has much more research to support information processing than the latter. Also, lower-level reading skills (e.g., word-level meaning) are closely associated with linguistic competence. Furthermore, L2 research is largely focused on product (i.e., are learners able to answer questions about texts) rather than process (i.e., how learners interact with texts). Thus, Devitt advocates a pedagogical approach in which learners first interact with words from an authentic text to create schemata for new vocabulary, then create their own stories with the words, put a scrambled version of the story in an ideal order, read the actual text, edit their own stories by interpreting some of the grammar from the actual text, and elaborate the text that they received. Though this approach needs more research in an L2 context, it is highly supported by approaches to literacy development.

Dings, A. (2014). Interactional competence and the development of alignment activity. *The Modern Language Journal*, 98(3), 742–756.

Highlights:

1. *He and Young's (1998) interactional competence is fundamental to understanding language acquisition in that it attempts to understand how meaning is co-constructed and how participants in interactions collaboratively manage communication.*
2. *The analysis of alignment activities ("ways in which interlocutors demonstrate their intersubjectivity, by showing each other that they are understanding each other and being understood" (p. 744) are likely to aid understanding interactions in target languages.*

Dings attempts to support previous research that promotes interactional competence (IC) as a competence—in addition to Canale & Swain's (1980) grammatical competence, sociocultural competence, discourse competence, and strategic competence—that L2 learners must develop. Interactional competence is a unique competency in that it considers meaning a co-construction of interlocutors during an interaction. Learners must rely upon resources such as alignment activities. To study development of alignment activity over time, Dings followed a learner studying abroad for a year in Spain who engaged in practice conversations with a native speaker. As time progressed, the learner's various alignment activities allowed her to play an increasingly active role in the co-construction of meaning during interaction. Though this study features limited overarching applicability (there was only one subject whose selection for this study from a larger study abroad group is not fully explained), it does suggest a need for research in how meaning is co-constructed during an interaction. If learners do not develop interactional resources such as alignment activities, they will be unable to contribute fully to interactions. In a language classroom with multiple proficiency levels, this situation yields a dynamic in which learners of higher proficiency levels bear a heavier burden to construct meaning than their counterparts.

Fulcher, G. (1998). Widdowson's model of communicative competence and the testing of reading: An exploratory study. *System*, 26, 281–302.

Highlights:

1. *Empirical support for the highest level of processing in Widdowson's three-level model with language testing,³¹ but no support for other levels.*
2. *Though reading is receptive, it is also somewhat interactional.*

Fulcher attempts to validate Widdowson's model of communicative competence empirically. Such attempts are rare in language acquisition literature; few theorists seek empirical validation, posing a threat to practitioners who might be implementing work that is not practically sound. Fulcher attempted to validate Widdowson's model by creating a test for secondary learners (Mage = 16.6 years) with systematic, schematic, and procedural levels. Students were practiced readers (those who engaged in education in the L2 of English since the age of 11 or before) or unpracticed readers (those who had not formally taken English courses since the age of 11 but did participate in eight hours of English tuition per week). Results showed that one cannot adequately tap into linguistic knowledge with word associations as the researchers attempted to do because linguistic knowledge is complex and difficult to assess correctly. Additionally, given the correlations that emerged from the two test groups, no evidence was found to support the schematic level, but

³¹ Systematic level (linguistic competence); schematic level (ability to use language); procedural level (actual performance).

this finding could have stemmed from the difficulty in defining capacity for performance. Finally, there was evidence to support the procedural level of competence. The article asserts that though reading is receptive, it is somewhat interactional. For example, textual organization creates a relation between texts and readers; therefore, it is important to understand how readers can interpret the “illocutionary force” (p. 285) of a text.

Gass, S., & Selinker, L. (2008). Input, interaction, and output. In *Second Language Acquisition: An Introductory Course* (3rd ed., pp. 304–367). New York, NY: Routledge.

Highlights:

1. *It is overly simplistic to assume that second language use does not vary from first language use; one cannot simply plug the necessary grammar and vocabulary into any communicative context in order to engage in the L2 appropriately.*
2. *Input and interaction help L2 learners to understand social contexts and participate appropriately in the co-construction of meaning according to context.*
3. *Output should be contextualized as essential to the process of learning, as well as a way to prove that learning has occurred.*

In this book chapter, Gass and Selinker explore how input informs learners not only of appropriate syntactic and lexical constructions in the L2, but also of how social contexts inform L2 utterances and interactions. Krashen’s Input Hypothesis (1982) is central to understanding how input informs language acquisition and language learning; although it is limited, in that different “levels” of language learning are not defined by the theorist, Krashen contributed that comprehensible input (input at $i + 1$, or rather input slightly above a learner’s ability to understand language) would help learners to acquire language. In other words, a learner who is completely new to a world language should receive basic input that is indicative of a low-level novice learner (words and syntactic chunks) because those elements of input are above what the learner can produce, but not so far above that the learner cannot comprehend with support. Similarly, a learner who functions at a high intermediate level should receive input that is more indicative of what an advanced learner would be able to produce (e.g., narration in multiple time frames).

The authors provide an in-depth view of language learning being attributable not just to input, but also output and the feedback that results from language interaction. Feedback has the potential to indicate to learners when they have contributed non-target-like utterances and to clarify misunderstandings. An important distinction is that this information contributes to comprehension of language but not necessarily acquisition, since acquisition indicates internalized, permanent change in L2 abilities. Last, this chapter contributes that even though output serves as proof of language acquisition and learning, it is essential to the learning process. Essentially, learners learn while creating with language.

Hadley, A. O. (2008). On knowing a language. In *Teaching Language in Context* (3rd ed., pp. 1–50). University of Illinois: Urbana-Champaign.

Highlights:

1. *Different competences (particularly those proposed in the Canale and Swain (1980) framework for communicative competence) are considered regarding how to conceptualize language proficiency as “a wide range of abilities that must be described in graduated fashion in order to be meaningful” (p. 9).*
2. *The ACTFL conceptualization of proficiency is evaluated for its ability to exemplify how*

communicative competence might develop in learners, and though there is little empirical evidence to support its assertions, there is a clear focus on using language for meaning making in ways that are appropriate to context.

Hadley considers the Canale and Swain (1980) framework and all the competences that are associated with building language proficiency including linguistic competence, sociolinguistic competence, discourse competence, and strategic competence. After having laid a framework for describing language proficiency, the author describes the move within the United States to proficiency-based guidelines exemplified by ACTFL and attempts that ACTFL underwent to assess proficiency. Next, Hadley contributes the global stages of linguistic acquisition defined by Ellis (1985). Finally, the author offers needed clarifications about language proficiency, including that it is not a theory of SLA, method of teaching, syllabus or curricular design, and it does not imply a focus on grammar and linguistic errors. However, considerations of language proficiency do and should affect language teaching, language assessment, and curricular design.

Harklau, L. (2002). The role of writing in classroom second language acquisition. *Journal of Second Language Writing*, 11, 329–350.

Highlights:

- 1. Due to interactionist approaches to language learning, language variation, and language socialization, it is just as important that learners learn how to write in a second language and learn through writing in a second language.*
- 2. Even though much of SLA research and literature supports language acquisition through face-to-face interactions (or computer-mediated synchronous interactions), observations of an English as a second language classroom showed written texts and output to be more consistent, more varied, and more attributable to student learning.*

Though SLA literature tends to favor face-to-face interactions and other forms of synchronous communication, the author references that U.S. high school classrooms use multiple modes of communication (i.e., visual, literate, and oral/aural) to reinforce the communication of ideas. This idea is reinforced by Cope and Kalantzis' (2000) "multiliteracies"; in order for learners to negotiate meaning, particularly in light of the breadth of media texts available presently, learners need to be able to consider multiple modes of communication simultaneously. In order for a learner to be proficient in one such environment, he or she has to manage "a wide range of academic and interpersonal registers in both face-to-face modalities and more formal academic modalities" (p. 388). In doing so, learners will be able to learn language appropriate to spoken contexts and appropriate to written contexts instead of favoring one. To ensure written texts and input receive adequate attention, the author advocates that writing be treated as a vehicle to learn a second language rather than as an SLA goal.

Hyland, K. (2007). Genre pedagogy: Language, literacy, and L2 writing instruction. *Journal of Second Language Writing*, 16, 148–164.

Highlights:

- 1. Employing a genre pedagogy in the classroom in lieu of focusing solely on grammar and distinct communication processes enables learners to not only write texts that are appropriate to academic contexts but also to write texts that will help them in their everyday lives in the target language. Additionally, it is likely that such an approach improves their reading, as they will come to know what to expect in different genres of writing.*
- 2. The key principles of genre-based writing instruction are 1) Writing is a social activity; 2)*

Learning to write is needs-oriented; 3) Learning to write requires explicit outcomes and expectations; 4) Learning to write is a social activity; and 5) Learning to write involves learning to use language.

Hyland explores the benefits of genre pedagogy and how to implement one such pedagogy in the second language classroom. Hyland begins by outlining that teachers must prepare students for writing in a variety of contexts—both inside and outside the classroom. One can achieve this goal by using a genre-based pedagogy in which learners explore the context of target text types in order to employ proper formatting, voice, and register. To ignore these important contexts is to ignore the social meaning of texts. Freeform writing does improve fluency, but it does not prepare learners to participate in a society in which the target language is spoken. In order to prepare learners to participate socially with writing, they need to explore purpose, context, and audience. They also need to focus on their relative writing needs in certain situations. Such instruction is successfully supported when teachers explain exactly what is being studied and why. Furthermore, teachers should link what students already know to new contexts. Finally, teachers should incorporate direct language instruction, but principally as it pertains to a specific genre/purpose. For example, if a text type requires persuasion, learners should learn and review structures associated with persuasion in the target language.

Knez'ic', D., Elbers, E., Wubbles, T., & Haajer, M. (2013). Teacher's education in Socratic Dialogue: Some effects on teacher-learner interaction. *The Modern Language Journal*, 97(2), 490–505.

Highlights:

- 1. For L2 learners to participate actively in the co-construction of meaning in the classroom, educators must allow them to contribute to the co-construction of knowledge by asking follow-up questions, checking for understanding, paying attention to language, reformulating learners' utterances, encouraging learners to talk, providing feedback on both content and language, and giving sufficient wait time to learners after posing questions.*
- 2. Training in Socratic Dialogue proved beneficial to teachers dealing with L2 learners. Teachers in the experimental group began to work with students to co-construct knowledge rather than presenting it directly. They invited learners to contribute, ensured accurate wait times, and paid more attention to potentially difficult language three months after training—even though they were not instructed directly to use Socratic Dialogue when mitigating Teacher-Learner Dialogues (TLDs).*

In this study, the authors investigated whether training in Socratic Dialogues would enhance scaffolding of TLDs for advanced second language learners. Three months after undergoing training in Socratic Dialogues, teachers in the experimental group were able to provide interactional scaffolding for learners, thereby empowering them to co-construct meaning and knowledge in the classroom. These scaffolding measures included asking open-ended questions, asking follow-up questions, maintaining a slow pace for dialogue, allowing learners adequate time to express themselves, ensuring that learners spoke for more time than the teacher, inviting learners to contribute to the discussion, paying attention to potentially difficult language, and checking to see if the teacher understood the learner's meaning. Teachers paid attention to difficult language, even though they did not discuss L2 acquisition in their training. Results are encouraging for L2 acquisition courses in which learners engage in concept-based and cross-curricular approaches.

Lotherington, H., & Jenson, J. (2011). Teaching multimodal and digital literacy in L2 settings: New literacies, new basics, new pedagogies. *Annual Review of Applied Linguistics*, 31, 226–246.

Highlights:

1. *Multimodality, though present in everyday interactions such as face-to-face conversations, is increasingly prevalent in the lives of individuals given the ubiquity of digital texts. This scope requires that educators adapt their practices to use, evaluate, and produce multimodal texts in the classroom.*
2. *Digital technologies allow learners to engage and produce texts within the classroom and beyond.*

Lotherington and Jenson provide insight as to how the proliferation of digital media has affected the language acquisition classroom. Most obviously, learners are confronted with needing to engage with multimodal texts and to produce them. This engenders new literacy awareness (such as the realization that websites read more like images than like text) and serves to destroy the confines of the traditional classroom. Indeed, learners can acquire language through play in digital games and can use digital devices to move from the classroom to interpret language and to create with it. However, such a shift does necessitate that educators receive training on how to incorporate technology into the classroom, as many teachers are hesitant to embrace a shift to new texts for learning about which they know little.

Mayer, R. E., Lee, H., & Peebles, A. (2014). Multimedia learning in a second language: A cognitive load perspective. *Applied Cognitive Psychology*, 28(5), 653–660.

Highlights:

1. *This study was conducted in a sheltered learning environment to test the addition of redundant media or language on comprehension.*
2. *Adding redundant video to an audio lecture for students learning content in their second language improved comprehension.*
3. *Adding redundant captions to a video for students learning content in their second language did not improve comprehension (the solution to learning in a second language is not to use subtitles).*

This study consisted of two experiments on students learning content in sheltered classrooms to determine if adding redundant graphics to an audio lecture (Experiment 1) or adding redundant captions to a narrated video lecture (Experiment 2) improved comprehension. Researchers were also interested in seeing if the second scenario demanded so much processing of students that they didn't experience the comprehension benefits. They found in the first experiment that students did show increased comprehension with the addition of the graphics, possibly because it helped clarify meaning of words, though students did not report more effort and less difficulty. In the second experiment, there was no difference between the subtitle group and no subtitle group, which is possibly explained by the split attention effect that the learners experience when listening, watching, and reading. These results should be taken into account when educators prepare learners to work with multimodal texts; learners with lower proficiency levels may experience cognitive overload when trying to simultaneously process two stimuli.

Sylvén, L. K., & Sundqvist, P. (2012). Gaming as extramural English L2 learning and L2 proficiency among young learners. *ReCall*, 24(3), 302–321.

Highlights

1. *Playing digital games at an early age (11–12) in the target language can be beneficial to language acquisition. This is likely attributable to the quantity of L2 input and “scaffolded interaction in the L2” (p. 302).*
2. *Frequent gamers had the highest scores on national reading and listening proficiency tests in Sweden, followed by moderate gamers, and then by nongamers. Additionally, frequent gamers scored higher than their counterparts on vocabulary tests.*

In this Swedish study, the authors examined relations between extramural gaming and L2 acquisition among 11–12 year olds. Participants completed questionnaires and language diaries that documented their engagement in the L2 (i.e., English) outside of class. The study found that self-reported frequent gamers claimed to have learned the L2 mainly outside of school and reported themselves to be infrequent readers of the L1. Additionally, frequent gamers had higher mean scores on a vocabulary test and national reading and listening comprehension tests than their peers. *This study is correlational; causation cannot be claimed.* Results point to a promising trend regarding the benefits of digital games for L2 acquisition, though the study has some limitations: (a) gaming habits were self-reported (though the questionnaire data tended to corroborate the diary data), raising questions about correlations between proficiency and gaming habits; (b) it is unclear if more frequent gamers were frequent gamers because of their relatively higher levels of proficiency or if their proficiency levels were higher because of increased gaming. Additionally, researchers did not control for prior knowledge of English and preferred learning style. Most participants were from relatively high socioeconomic backgrounds, thwarting generalizability.

Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *The Modern Language Journal*, 82(3), 320–337.

Highlights:

1. *The study provided support for a theoretical orientation toward viewing dialogue as a means of communication and a cognitive tool.*
2. *Collaborative dialogue is a useful concept for understanding L2 learning.*

Swain and Lapkin replicated a study by Goss et al. (1994) that observed how students worked together to decide if something was grammatically correct or not, this time observing collaborative dialogue between students and how it resulted in their understanding of an L2. Participants—Grade 8 students who had been in an early French immersion program since kindergarten—received a pretest, collaborative tasks, and a posttest. In recorded conversations, researchers observed students’ negotiation of meanings, attempts with new forms of the L2 with partners, and effects on posttest results. Conversations with their partners about the L2 associated with items marked incorrectly on the pretest being corrected on the posttest.

Weigle, S. C. (2004). Integrating reading and writing in a competency test for non-native speakers of English. *Assessing Writing*, 9, 27–55.

Highlights:

1. *The study examined and compared the results of an old literacy placement test, Alternate Regents Test, and a new version of that test that included an integrated reading and writing section. The more integrative test provided more reliable results for placement and contributed to a positive washback effect for instructors preparing learners to take the placement exam.*
2. *The importance of developing new forms of tests to ensure validity and fairness.*

In this study, Weigle provided insights both about the potential of a reading-writing connection and potential downfalls of high-stakes essay examinations for nonnative speakers of English. The researcher evaluates a reading/writing examination created by a committee of Applied Linguistics/English as a Second Language faculty. This test was meant to replace a state university's placement exam for nonnative English speakers. Prior to the decision to revise the exam, faculty expressed concerns that the previous test, which was 20 years old, did not accurately assess learners' proficiency levels (e.g., writing prompts devoid of context; limited consideration of audience or purpose outside completion of the task itself). In the new exam, learners no longer write a one-time timed essay; they engage in a multiple-choice reading section and integrated reading/writing section for which they respond to two texts with differing perspectives about an issue of global significance. Student responses are evaluated for content, organization, language accuracy, and language range and complexity. After two years of delivery, more learners were placed appropriately. Interrater agreement increased from 65% on the former test to 95%. Also, educators teaching the prep course for the reading/writing exam focused on skills that are "arguably more relevant to other academic courses" (p. 44).

Yuksel, D., & Inan, B. (2014). The effects of communication mode on negotiation of meaning and its noticing. *ReCALL*, 26(3), 333–354.

Highlight:

1. *As described in this article, face-to-face (F2F) communication facilitates more negotiation of meaning (NoM) than synchronous computer-mediated communication (SCMC), but SCMS does increase more instances of noticing.*

The crux of the Interaction Hypothesis is that NoM in the L2 facilitates language acquisition. This hypothesis rests on the observation that NoM requires L2 learners to consider feedback regarding miscommunication and appropriate output. Essentially, when communication breaks down, learners are afforded the opportunity to consider input and modify output, thereby acquiring more target-like language. To consider how NoM occurs in F2F communication and in SCMC environments, researchers worked with 64 young adult English learners and had them engage in jigsaw tasks. Each learner engaged in one F2F jigsaw and one SCMC-mediated jigsaw. Researchers used a stimulated recall technique (participants received scripts or film recordings of their interactions and had to document what they were thinking about in every instance of negotiation of meaning) in order to gauge to what extent learners had noticed their errors. Results aligned with previous studies: F2F optimizes instances for NoM when compared to SCMC. Still, learners were more likely to notice and recall their errors in computer-mediated environments. Providing that noticing one's errors involves understanding, SCMC may help practitioners in providing learners with the appropriate feedback to produce uptake.

Zuengler, J., & Miller, E. (2006). Cognitive and sociocultural perspectives: Two parallel SLA worlds? *TESOL Quarterly*, 40(1), 35–58.

Highlights:

1. *There are competing, likely irreconcilable, ontological perspectives regarding the sociocultural and cognitive natures of SLA; however, this situation inspires robust discussions regarding language acquisition that aids the field.*
2. *Despite the competing natures of the perspectives, all agree that language learning is a cognitive process and that social factors must be considered when understanding (a) how target-like utterances are constructed and (b) how to construct target-like utterances.*

In this article, the authors provide an overview of contemporary SLA theory and tension between approaches that value the social traits of L2 acquisition over the cognitive traits and vice versa. From an ontological perspective, these differences are largely irreconcilable, though certain axioms emerge when considering them. For example, when pondering various frameworks that describe a progression of language learning and the cognitive development of learners, it is essential to consider Vygotskian sociocultural theory and its call that a learner work within his or her Zone of Proximal Development—the learner’s potential to acquire knowledge and skills with help and then ultimately, hopefully, without help. Another important discussion in the article is the contribution of Firth and Wagner (1997) and their conclusions that SLA research leading up to their publication focused too much on the individual (i.e., internalization of grammar and other processes) as a learner and not on the co-construction of meaning among individuals involved in the social traits of communication. Educators concerned with their learners’ sociocultural development must be equally concerned with preparing learners to communicate within and outside the classroom. Though theorists debate the degrees to which L2 acquisition is both cognitive and social, the presence of both factors is accepted.

III. Assessment

Bigelow, M., Delmas, R., Hansen, K., & Tarone, E. (2006). Literacy and the processing of oral recasts in SLA. *TESOL Quarterly*, 40(4), 665–689.

Highlight:

1. *Learners with low alphabetic literacy in both L1 and L2 recalled recasts with less success than more literate peers: findings from previous studies with respect to oral L2 processing of literate learners may not be applicable to other learners.*

To see if the benefit that literate learners experience by engaging in interaction in L2 through NoM and feedback applies to learners with low literacy levels, Bigelow et al. evaluated adult Somali immigrants with hopes that the strong oral tradition within Somali culture would have developed aural skills that transfer to L2 learning. Learners with lower literacy levels had difficulty in recalling recasts. Though the length of the recasts seemed to bear no impact on this ability, only learners with higher levels of proficiency could consistently and correctly recall recasts. This finding has implications for language learning contexts, particularly in countries with high transient populations, but lacks the replication size to make any of the findings generalizable. Still, practitioners should take care to place learners in the appropriate phase of language acquisition as low literacy in L1 may affect the speed of L2 proficiency development.

Egi, T. (2010). Uptake, modified output, and learner perceptions of recasts: Learner responses as language awareness. *The Modern Language Journal*, 94(1), 1–21.

Highlights:

1. This study analyzed the relationship between recasts, uptake, and output in a group of learners of Japanese.
2. Uptake is an indicator that learners have understood recasts as corrective feedback. In addition, error repair and modified output correlates with understanding corrective feedback and noticing discrepancies between interlanguage and L2.

In this study with learners of Japanese, Egi evaluated relations between recasts, uptake, and modified output. Learners completed classroom tasks while receiving recasts, then watched videos of their interactions and reported their thought processes at the time of each recast. Results showed that learners perceived recasts as corrective feedback more often when they responded to recasts, helping them engage in the critical repair of L2 utterances. In cases where learners repaired utterances, they were more likely able to report interlanguage/L2 discrepancies. In addition, in instances where learners made modifications to their output, they were more likely to report noticing the corrective recasts and identifying the interlanguage/L2 relation. When learners repeated the same mistake, they very rarely reported noticing the corrective nature of the recasts.

Frey, B., Schmitt, V. L., & Allen, J. (2012). Defining authentic classroom assessment. *Practical Assessment, Research, & Evaluation*, 17(2), 1–18.

Highlights:

1. Definitions of authentic assessment are varied: “real-world like” is not comprehensive.
2. Authentic assessment often includes many different characteristics with the goal of contrasting traditional assessment.
3. Authentic assessment refers to involving the student cognitively and with regard to interest, and developing or evaluating skills and abilities that have value beyond the assessment.

In this literature review, the authors analyzed the definitions of authentic assessment that scholars have proposed since 1988. Necessary characteristics of authentic assessment ranged from involving realistic tasks, collaboration, and self-assessment; being performance-based, cognitively complex, and formative in nature; and including multiple indicators of success and concrete expectations of mastery. Being realistic was mentioned in about half the literature—about the same as the use of multiple indicators. The authors decided that though being realistic is an important trait of authentic assessment, it is not a sufficient descriptor. Thus, the authors ignore some definitions used in favor of ones that signified the “unique worth of the approach” used in assessment (p. 14). These definitions include involving students cognitively, including student voice and interests, and developing or evaluating skills and abilities of value beyond the assessment in question.

Mackey, A. (2006). Feedback, noticing, and instructed second language learning. *Applied Linguistics*, 27(3), 405–430.

Highlights:

1. Though more research needs to be conducted, researchers posit that L2 interaction facilitates language learning by providing interlocutors with feedback regarding non-target-like utterances that they can hopefully rectify to make more target-like.
2. Learner reports of noticing certain target forms (past tense, plural words, and question formation) correlated positively with their L2 development for those forms.

No study shows learning to be clearly dependent on noticing (i.e., learners may develop without reporting having noticed and learners may say they noticed but never actually improve their L2 utterances). Still, in this study Mackey found a correlation between an L2 learner noticing target language forms and the learner learning those target language forms. The study featured two groups of students who were involved in the same input and output activity (a pop culture game show). The control group received no input to facilitate their noticing of target structures, unlike the experimental group. Each group took notes during the input and output activity and completed journals to evaluate their progression with regard to the target structures. Learners who received input to facilitate their noticing of the target structures showed more growth with respect to those target structures than their peers. Learners showed much more internalization of rules regarding question formation than those regarding past tense verbs. It is possible that the documents in which learners took notes directed them to consider those structures more closely, though it is unclear what specific characteristics learners were asked to observe (i.e., the lone example addressed pronunciation specifically, no target features). One also wonders why speech acts (e.g., requests) were not observed in lieu of disparate structures. Such an approach would more closely align findings with the language functions advocated by ACTFL and the CEFR.

Plakans, L. (2009). The role of reading strategies in integrated L2 writing tasks. *Journal of English for Academic Purposes*, 8, 252–256.

Highlights:

- 1. Though much research supports a link between successful readers and successful writers, less is known regarding what reading strategies successful writers employ and how those strategies affect written output. The common practice of integrating reading and writing into a single assessment necessitates such research.*
- 2. English learners at U.S. universities (n = 12) were asked to write an argumentative essay using a text regarding cultural borrowing and technology as support. Bottom-up reading strategies such as breaking down word-level meaning were the strategies most commonly used. Scanning, skimming, and responding to ideas with personal opinions were also common. These global reading strategies were the most common among the most proficient language writers; bottom-up strategies were generally the only strategies that less proficient writers employed.*

In this study, Plakans considered how writers facing integrated assessments used different reading strategies, examining how those strategies correlated with written output. Though the most common strategies used by all 12 participants were word-level processing strategies, this finding can be attributed to the vast quantity of individual words in a text. High-scoring writers used the largest quantity of high-level reading strategies (e.g., skimming and making personal connections with the texts) but, given the small sample size, these results are not generalizable. Also, the study cannot facilitate any causal inferences between reading and writing skills. Still, the data do point to a need for further study of a taxonomy of reading strategies (e.g., mining, reading comprehension, and goal setting) and how those strategies influence written output.

Mantero, M. (2002). Evaluating classroom communication: In support of emergent and authentic frameworks in second language assessment. *Practical Assessment, Research, & Evaluation, 8(8), 1–7.*

Highlights:

1. *Though sociocultural theory predominates much of SLA pedagogy, many testing methods do not reflect this shift, focusing instead on discrete grammar and vocabulary knowledge. Such a limited approach ignores contexts and cultures, influences that critically affect communication. Furthermore, an approach based on a priori grammar does not measure the discursive capabilities of a learner, nor does it allow a learner to reach beyond the Zone of Actual Development.*
2. *Wiggins' authentic assessment is a good approach for language learning because it requires a mirroring of real-life communication. To find meaning, learners need feedback about their respective and potential language proficiencies.*

Mantero examines sociocultural theory and pedagogy in world language classrooms, recommending how assessments should be developed to marry pedagogy and assessment. Following Vygotsky, whose theories contextualized communicative tasks as leading to the acquisition of language, the paper clarifies a language teacher's goal as moving learners from dialogue to discourse. Teachers should have learners exchange utterances, all the while negotiating meaning; teachers should scaffold students into being able to include their own ideas and opinions so that discussions become discursive. Scaffolding is essential so that a student's Zone of Proximal Development becomes a Zone of Actual Development. Instead, traditional language assessments focus largely on a priori grammar. Thus, they indicate linguistic knowledge and not the capacity to use that knowledge. In essence, a student would never have had to move beyond the Zone of Actual Development to succeed on such an assessment. An advocated approach, therefore, is to design communicative assessments, or what Wiggins (1989) would call authentic assessments. Ideally, these assessments would be discursive and would incorporate authentic communication tasks in authentic cultural contexts, even if those contexts have to be simulated in the classroom.

Pellet, S. (2012). Wikis for building content knowledge in the foreign language classroom. *CALICO Journal, 29(2), 224–248.*

Highlights:

1. *Wikis engaged students and provided an authentic, meaningful, active experience.*
2. *Collaborative design can enhance student-centered computer applications outside and inside the classroom.*
3. *Technology-assisted learning creates new roles for teachers: expert and a guide.*

In a qualitative case study, Pellet shows two wikis used with French learners and how learners engaged with wikis can work collaboratively. Students used wikis for knowledge building and presentation, collaboration, and content and language learning. At the end of the semester, students provided feedback about the wiki. All students reported that the main payoff was reviewing for tests, but they also listed interacting in small groups, developing critical and analytical skills, connecting with French users, taking ownership of technology, collaborating, and enjoying the experience as other successes. The instructor reported authentic student ownership as a success. There was also a focus on process and independent learning; both collaboration and cooperation took place.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *The Phi Delta Kappan*, 70(9), 703–713.

Highlights:

1. *In order for assessments to be an integral driver of student performance, assessments should replicate authentic problems that people encounter in the real world.*
2. *Authentic assessments give practitioners insights regarding the mental processing of students in lieu of their short-term recall.*
3. *Authentic tests are “contextualized, complex intellectual challenges, not fragmented and static bits of tasks” (p. 711).*

Though this article falls outside the time frame of this study, it is foundational and employed by IB in the development of its materials. Therefore, it has been included for its relevance. Wiggins was revolutionary for advocating the idea that learners engaging in a subject area should be amateur practitioners of those subject areas and attempt the problems and tasks that one would within those fields. Such an approach allows practitioners to evaluate how learners process information and allows learners considerable voice in how they are assessed (i.e., they can ask for clarification of topics, etc.). Strong relations exist between this approach and NoM tasks, in which learners receive the chance to repair their non-target-like L2 utterances. Authentic assessment is particularly important for SLA because researchers have shown learners to acquire language through communicative tasks that are contextualized and meaningful.

Wiggins, G. (2011). Moving to modern assessments. *The Phi Delta Kappan*, 92(7), 63.

Highlights:

1. *Wiggins clarifies that “authenticity” is achieved in assessment less through task creation and the “realism of the setting.” Realism considers the context of the assessment.*
2. *Wiggins extends his original discussion of authentic assessment to include the notion that learners can transfer what they have learned to new and unfamiliar contexts.*

Though this article is not peer-reviewed, it is included here as an important follow-up to Wiggins’s seminal article. After two decades of work with authentic assessment, Wiggins realized that context is incredibly important when constructing an assessment. Skills that learners build within a context will hopefully help them transfer meaning to new contexts, situations that are common to language learning. It is practitioners’ duty to help learners to understand communicative expectations (e.g., power dynamics, language register, and communicative goals) within a given context, so that students might be empowered to function within other contexts.

APPENDIX B: FULL-LENGTH MYP LITERATURE REVIEW

Research Question 1: *How effectively does the MYP language acquisition course conceptualize and describe a progression of learning?*

Purpose

The purpose of this literature review is to analyze how effectively the International Baccalaureate's (IB) Middle Years Programme *Language Acquisition Guide* (MYP *Guide*) conceptualizes and describes a progression of learning. We address this topic with a comprehensive literature review to create a picture of learners' progressions in language acquisition. Second-language acquisition (SLA) is interdisciplinary, complex, and difficult to describe (Gass & Selinker, 2008). This literature review serves to bring some cohesion to the considerably broad scope of research on language acquisition.

Overview

To ensure that the most relevant resources informed our review, we included sources that met three criteria: they (a) were peer-reviewed; (b) addressed listening and speaking, viewing and interpreting, reading comprehension, and writing; and (c) offered insights into how effectively IB assessment criteria map onto its progressions of language learning in the MYP *Guide* and/or how the developmental progression specified in the MYP *Guide* links with the language acquisition development progressions of other IB programmes (i.e., the Primary Years and Diploma Programmes). Whenever possible, we included studies that directly address language learning processes of MYP-age learners (i.e., 10–16 years). However, given that the IB devised the MYP language acquisition global proficiency table using the Council of Europe's (2002) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (CEFR) and the American Council on the Teaching of Foreign Languages' *Standards for Foreign Language Learning in the 21st Century* (ACTFL; 2011)—and neither of these documents are specific to MYP-age learners—it was necessary to expand our literature review to include studies involving language learners of all ages. Furthermore, MYP ages represent a bracket that is under-represented in SLA literature. To contend with this deficiency and provide a robust review of relevant research, we considered all literature published since 1997 that met our criteria; we emphasized studies from the past 10 years.

After we conducted a cursory review of qualifying studies, clear areas of alignment emerged regarding the MYP *Guide*. Areas of alignment include: (a) the guide's phases of language acquisition; (b) the relation between the four language continuums (listening and speaking, viewing and interpreting, reading comprehension, and writing) and the four language processes (comprehending spoken and visual text, comprehending written and visual text, communicating in response to spoken and/or written and/or visual text, and using language in spoken and/or written form); and (c) consideration of assessment task design. We close the literature review with recommendations for future iterations of the MYP *Guide*.

Method

To investigate the appropriateness of the phases of language acquisition, we sought to evaluate the MYP *Guide*'s use of CEFR and ACTFL standards. To parse this research and its relation to

the phases in the *MYP Guide*, we searched both educational and language acquisition journals for any articles that addressed phases or stages in language learning. The search terms “phases in L2 acquisition,” “phases in SLA,” “phases in language learning,” “stages in L2 acquisition,” “stages in SLA,” and “stages in language learning” did not yield many results, possibly because contemporary literature negates the assumption that language acquisition follows a linear progression (Westhoff, 2007). To find more relevant sources, we searched the same journals using the terms “Validity ACTFL” and “Validity CEFR,” which yielded a plethora of results largely focused on the history of establishing the standards, the lack of empirical research to support the standards, and their degree of proliferation. The dual recognition that there is a lack of empirical research to support the standards and that their high degree of proliferation perpetuates their existence proved insightful when analyzing the *MYP Guide*’s phases of language acquisition.

Next, we investigated the four language processes and their relations with the language continuums. Immediately, we noticed that neither the language processes nor the continuums align directly with CEFR or ACTFL specifications. Given the degree of proliferation of CEFR and ACFTL, research tends to follow the standard sets by delineating macrolevel language skills. For example, interpersonal communication, an ACTFL mode, and the assessment of it, can be found commonly in research literature. However, MYP’s Assessment Criterion C: Communicating in response to spoken, written, and/or visual text does not align directly with research regarding interpersonal communication due to the source text requirement. Such misalignment is compounded by MYP’s subject-specific guidance that this criterion may be assessed with a presentational task (see *MYP Guide*, p. 43). Furthermore, the MYP language continuums do not account specifically for written synchronous communication (e.g., texting). Though it is possible that such interpersonal communication could be accounted for in both the Reading and Writing continuums, such an approach still indicates a disparate treatment of macrolevel communication between an internationally recognized conceptualization of communication and the *MYP Guide*.

To avoid complications stemming from this misalignment, we chose to employ the heuristic technique of evaluating the language processes in the *MYP Guide* for their potential to engage learners in meaningful communication, that is, communication that will produce and be indicative of language acquisition. Accordingly, we searched education and language acquisition journals and texts to identify the most important characteristics of meaningful communication. This search necessitated that we expand our temporal inclusion parameter to consider Canale and Swain’s (1980) framework³² for communicative competence since it is foundational in contemporary language acquisition literature. This search also led us to consider literature pertaining to effective input, comprehension and interaction, and output. Then, we evaluated the theoretical underpinnings of the *MYP Guide*’s focus on visual interpretation in combination with reading, writing, listening, and speaking. Our evaluation included considerations related to Byram’s (1997) model for intercultural competence, contextualized communication, and literature regarding multiliteracies and computer-mediated communication.

Finally, we were able to consider the assessment task design that the *MYP Guide* describes. Given IB’s focus on authentic, student-centered assessment, we expanded our temporal inclusion parameter by seven years to include (a) Wiggins’ original article (1989) on the need for authentic assessment and (b) an opinion piece by Wiggins (2011) that reflects how the principles for authentic

³² Canale (1983) updated the framework, and the updated version informs the literature review and is the object of reference for mentions of Canale & Swain’s framework unless otherwise noted.

assessment have influenced pedagogical approaches to assessment over the past 20 years.³³ We built upon Wiggins' research with more contemporary views of authentic assessment to evaluate how context is embedded within the assessed language processes. Given that the exploration of context is a central axis within genre-based language acquisition pedagogies (an approach in which learners consider how to structure communication through analysis of the contextual variables that are pertinent to each text type discussed), we considered education and language acquisition literature regarding that theme.

We close our review of assessment task design by considering the preoccupation that emerges within both second language acquisition literature and education literature regarding comprehensibility of feedback when multiple processes are being assessed simultaneously. In the language acquisition context, this situation arises within communicative tasks that involve the concurrent use of multiple modes of communication. We made this decision due to the observation that Assessment Criteria A, B, C, and, depending on design, perhaps D, all involve some sort of input and output. The communicative focus of these assessments led us to revisit literature regarding meaningful communication to explore links between formative and summative assessments. Generally, the literature emphasizes the importance of using communicative tasks formatively rather than summatively, a consideration that is crucial for evaluations centered on structuring classrooms to build proficiency (i.e., learning to use language in lieu of simply learning about language), a clear objective of the *MYP Guide*.

Phases of Language Acquisition

Defining Phases of Language Acquisition

Beginning on page 11 of the *MYP Guide*, IB divides the progression of language learning into six phases. These phases link most closely to other conceptualizations of language learner progression such as those purported by CEFR and ACTFL. They *MYP Guide* and the two international standard sets put forth phases in which language and the contexts in which it is used become increasingly complex as learners progress. In Table A1, we highlight these progressions with regard to the complexity of spoken or written interactions in the target language. Due to the differing ways in which the standards sets describe their various approaches, we chose the closest approximation of alignment for Table A1. The alignment of these conceptualizations is not the focus of this literature review. We include it here as an analytical example only.

As Hadley (2008) comments, "Proficiency is not a monolithic concept representing an amorphous ideal that students rarely attain; rather, it is comprised of a whole range of abilities that must be described in a graduated fashion in order to be meaningful" (p. 8–9). The phases of language acquisition denoted by the *MYP Guide* embody this sentiment. As shown in Table A1, not only does language complexity increase as learners move from Phase 1 forward, but topics and contexts also become more broad and varied. Communication acts, not syntactic and lexical accuracy, are the focus. This progression reflects the contemporary proficiency movement in language acquisition that rejects previous approaches (e.g., grammar-translation). It follows then that the progression of communication described in the *MYP Guide* for the phases of communication is logical, though it is only corroborated by the experience of practitioners. The alignment of the *MYP Guide* with international standard sets is clear and is featured in Table A1.

³³ Wiggins' writings are the only sources that inform our literature review that exist outside the original parameters set forth in our specified methods and were included given the reasons articulated in this discussion.

Table A1. Alignment of MYP Criterion C, CEFR Qualitative Aspects of Spoken Language Use, and ACTFL Performance Descriptors

	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6
MYP LA Criterion C	Use basic phrases to communicate ideas, feelings and information on a variety of aspects of everyday topic.	Use phrases to communicate ideas, feelings, and information in familiar situations.	Express ideas and feelings, and communicate information in familiar and some unfamiliar situations.	Express ideas and feelings, and communicate information in simple and complex texts in familiar and unfamiliar situations.	Express ideas, opinions, and feelings, and communicate information in a wide range of situations.	Express a wide range of ideas, opinions and feelings, and communicate information in a wide range of social and academic contexts.
	Breakthrough	Waystage	Threshold	Vantage	Effective operational proficiency	Mastery
CEFR Qualitative aspects of spoken language use	Can ask and answer questions about personal details. Can interact in a simple way but is totally dependent on repetition, rephrasing, and repair.	Can answer questions and respond to simple statements can indicate when he/ she is following but is rarely able to understand enough to keep conversation going of his/her own accord.	Can initiate, maintain and close simple face-to-face conversation on topics that are familiar or of personal interest. Can repeat back part of what someone has said to confirm mutual understanding.	Can initiate discourse, take his/ her turn when appropriate and end conversation when he/ she needs to, though he/ she may not always do this elegantly. Can help the discussion along familiar ground confirming comprehension, inviting others in, etc.	Can select suitable phrase from a readily available range of discourse functions to preface his remarks in order to get or to keep the floor and to relate his/ her own contributions skillfully to those of other speakers.	Can interact with ease and skill, picking up and using nonverbal and intonational cues apparently effortlessly. Can interweave his/ her contribution into the joint discourse with fully natural turntaking, referencing, allusion making, etc.
ACTFL (2012) Performance Descriptors	Novice	Intermediate	Advanced			
	Expresses self in conversations on very familiar topics using a variety of words, phrases, simple sentences, and questions that have been highly practiced and memorized	Expresses self and participates in conversations on familiar topics using sentences and series of sentences. Handles short interactions in everyday situations by asking and answering a variety of questions. Can communicate about self, others, and everyday life.	Expresses self fully to maintain conversations on familiar topics and new concrete social, academic, and work-related topics. Can communicate in paragraph-length conversations about events with detail and organization. Confidently handles situations with an unexpected complication. Shares point of view in discussions.			

Limitations of Phases in Language Acquisition: Empirical Evidence

Unfortunately, little empirical evidence serves to support these various conceptualizations of the progression of language learning (Hulstijn, 2007; Hulstijn et al., 2011; Liskin-Gasparro, 2003; Little, 2007; North, 2007), which produces circularity in the research. For instance, Liskin-Gasparro (2003) critiques ACTFL's guidelines: "I am an Intermediate-High speaker of French because I can perform the tasks associated with the Intermediate-High level with the requisite accuracy; and these tasks and this degree of accuracy are associated with the Intermediate-High level because people . . . who are rated Intermediate-High, are able to do them" (p. 485). In other words, the guidelines must be interpreted with caution because the guidelines themselves provide their own justifications. Clifford and Cox (2013) did find evidence to support that reading skills are hierarchical in English, Spanish, and Chinese in a validation study of ACTFL's proficiency guidelines, but overall empirical evidence for the progression is limited in quantity and by various delineations within the frameworks. For example, the modes of communication defined by the frameworks are the focus of the research, thereby limiting the possibility that new or different conceptualizations of communicative modes and how those modes might progress will emerge. Thus, the phase-specific progression of objectives within the *MYP Guide* (p. 11–14) and of the language continuums (p. 28–35) might be corroborated by more empirical evidence if the parameters of contemporary research were defined differently. Indeed, it would be preferable that some sort of corpus of learner-produced evidence served as the foundation upon which the frameworks were developed in lieu of the corporate experience of various practitioners, which provides extant support (Hulstijn, 2007). A corpus-based approach would also be beneficial in ensuring that no critical language functions are omitted from student learning targets. For example, Leung and Lewkowicz's (2013) study of classroom interactions showed that nonnative speakers contributed to classroom conversations and discussions using language functions not specifically highlighted by the CEFR. In the absence of any such corpus or additional repository of evidence to support a shift in the way the field defines proficiency levels of language learners, there is perhaps no better vehicle to document the progression of language learning than experience.

Limitations of Phases in Language Acquisition: Complex Progression

The use of guidelines such as those provided by CEFR and ACTFL is further problematized by the lack of linearity in typical language acquisition, a process that is neither neat nor easy to describe (Westhoff, 2007). For example, it is possible that grammatical development occurs in stages and gradually (Smith & Truscott, 2005). Hulstijn et al. (2011) emphasize this trajectory: "The CEFR levels, in their present form, are neither based on empirical evidence taken from performance, nor on any theory in the fields of linguistics or verbal communication" (p. 204). Thus, though defining proficiency levels certainly simplifies any discussions one might have when discussing learners' abilities, the distinction between each level is somewhat arbitrary. Language acquisition theorists such as Stephen Krashen underscore the arbitrary nature of these levels by failing to define separate levels of knowledge (Gass & Selinker, 2008), rather only affirming that certain levels exist.

Limitations of Phases in Language Acquisition: Proliferation

The final critique of the CEFR and ACTFL guidelines is their degree of proliferation. Given their scope, it is unlikely that robust examples of alternative conceptualizations of language acquisition exist. Liskin-Gasparro (2003) postulates that the tentacles of ACTFL have such broad

reach because it positioned itself as an “agent of change” (p. 487) when the language proficiency movement became a national trend in the late 1970s and early 1980s in the United States. The same status is certainly applicable to the CEFR, a status that has yielded the alignment of teaching and assessment resources to these frameworks. As a result, creators of examinations, textbooks, and curriculum frameworks perpetuate CEFR (Alderson, 2007) and ACTFL. In other words, the proliferation of CEFR and ACTFL frameworks disincentives the creation of alternative teaching and assessment tools which could stimulate additional language acquisition insights.

Further Complications: Research Considerations

Though researchers call for more empirical evidence to support how language is acquired and to validate CEFR and ACTFL, several challenges thwart the collection of relevant data, particularly in classroom environments. First, communication is complex and multimodal. In any given conversation, interlocutors may be affected by myriad influences including their own cultural backgrounds, the communicative context, embedded power dynamics, and/or linguistic and grammatical competence. Spada (2005) discusses this interdependence of factors, saying that they are difficult to isolate for research purposes, thereby raising an epistemological concern for the validity of such research. Secondly, technical considerations including selection of practitioners and their training and relative biases affect such research. Finally, to sustain necessary relationships with practitioners, researchers must persuade them that the findings are applicable, which requires sustained efforts during dissemination (Spada, 2005). Taken in combination, factors in this section and its predecessors support IB’s decision to consider CEFR and ACTFL when creating its language acquisition phases, although the limitations of such an approach remain. The frameworks are globally recognized and proficiency-focused. Moreover, few alternatives exist.

Four Language Processes and the Blending of Continuums

The MYP *Guide* divides communication into three “areas” (oral, visual, and written). Four “continuums” represent these areas: (a) listening and speaking, (b) viewing and interpreting, (c) reading comprehension, and (d) writing. Taking into account that these areas of communication oftentimes interact (e.g., when watching a commercial on television, audiences interpret images alongside audio), IB disentangled and recombined these continuums into the four language processes/assessment criteria:

Criterion A: Comprehending spoken and visual text

Criterion B: Comprehending written and visual text

Criterion C: Communicating in response to spoken and/or written and/or visual text

Criterion D: Using language in spoken and/or written form.

In general, this approach provides a realistic model of language use. To explore this assertion, we consider how these language processes/assessment criteria relate to Canale and Swain’s (1980) framework for communicative competence and will discuss their potential to engage learners in effective language learning.

Canale and Swain (1980) proposed a framework for communicative competence that remains widely influential. They posit three components of communicative competence: (a) grammatical, (b) sociolinguistic, and (c) strategic. Canale (1983) added discourse competence later. In Table A2, we outline how engaging in IB’s assessment criteria develops each competence.

Table A2. Canale & Swain's Components of Communicative Competence Described and Related to IB Assessment Criteria

Communicative competence	Description	Relation to assessment criteria
Grammatical competence	A person's ability to use language appropriately and according to convention	<p>Criterion A: Analysis of source texts may influence learners' development of grammatical competence and may contribute to meaning making (Strand iii) / understanding of meaning (Strand i).</p> <p>Criterion B: Analysis of source texts may influence learners' development of grammatical competence and may contribute to meaning making (Strand iii) / understanding of meaning (Strand i).</p> <p>Criterion C: Grammatical competence influences a person's ability to respond in interactions (Strand i) and use language (Strand ii). It also influences one's ability to communicate ideas (Strand iii).</p> <p>Criterion D: Vocabulary, structures, and conventions are directly assessed in Strand i of this assessment criterion.</p>
Sociolinguistic competence	A person's ability to use language appropriately in a given communication context	<p>Criterion A: Analysis of features in a source text may influence learners' development of sociolinguistic competence, particularly with respect to conventions (Strand ii).</p> <p>Criterion B: Analysis of source texts may influence learners' development of sociolinguistic competence, particularly with respect to conventions (Strand ii).</p> <p>Criterion C: Sociolinguistic competence is developed through practice of Strands i, ii, and iv (responding appropriately, interacting using verbal and nonverbal language, and communicating with a sense of audience).</p> <p>Criterion D: Strand iii (use language to suit the context) requires sociolinguistic competence.</p>
Discourse competence	A person's ability to produce cohesive and complex thoughts	<p>Criterion A: Discourse competence can be developed by understanding conventions (Strand ii) and by producing a personal response to the text (Strand iii).</p> <p>Criterion B: Discourse competence can be developed by understanding conventions (Strand ii) and by producing a personal response to the text (Strand iii).</p> <p>Criterion C: Discourse competence can be portrayed through communication of ideas and thoughts (Strand iii).</p> <p>Criterion D: Discourse competence can be portrayed through the organization of communication (Strand ii).</p>
Strategic competence	A person's ability to use strategies to compensate for gaps in knowledge and misunderstandings	<p>Criterion A: Strategic competence is developed through learners' abilities to identify information (Strand i and to some extent Strands ii and iii) when confronted with not understanding all the source text and not knowing exactly how to say everything that is desired to be said but communicating the ideas anyway.</p> <p>Criterion B: Strategic competence is developed through learners' abilities to identify information (Strand i and to some extent Strands ii and iii) when confronted with not understanding all the source text and not knowing exactly how to say everything that is desired to be said but communicating the ideas anyway.</p> <p>Criterion C: Strategic competence is developed through learners' abilities to respond in interaction and convey and interpret messages with verbal and nonverbal language (Strands ii and iii) when confronted with misunderstandings or noncomprehension.</p> <p>Criterion D: Strategic competence is developed when learners communicate their ideas (Strand i) when confronted with linguistic limitations.</p>

In Table A2, we illustrate that the assessment criteria in the *MYP Guide* provide learners with the opportunity to prove meaningful language use; all assessment criteria have the potential to foment learners' development of communicative competences. Furthermore, the guide does not overemphasize grammatical competence. As Devitt (1997) outlines, this grammar comprises lower-level skills. When cultivating learners into competent communicators, practitioners must focus on developing the higher-level skills (e.g., sociolinguistic, discourse, and strategic) that are necessary for processing most authentic input.

Effective Input, Comprehension and Interaction, and Output

One of the ways to develop all language competences is to ensure that language learners encounter effective input, comprehension and interaction, and output (Gass & Selinker, 2008). In combination, they provide optimal situations for language learning. We discuss the potential of the language processes to foster input, comprehension and interaction, and output here.

Input

The *MYP Guide* features assessment criteria that allow for a myriad of text types from which to provide learners with input. For example, Criterion A articulates visual and spoken texts, Criterion B articulates visual and written texts, and Criterion C articulates communicating in response to a written and/or visual text. Though little guidance is provided to determine what these source texts might be, the open-ended nature gives practitioners the freedom to capitalize on a variety of texts. Allowing for such breadth increases the likelihood that learners might receive enough comprehensible input to grow their proficiencies. As Gass and Selinker (2008) outline, such input should fall within what Vygotsky called a learner's zone of proximal development. For instance, positive correlations between video game learning and increased proficiency among middle school students suggest that limiting source text variety might limit learners' opportunities to improve language proficiencies (Sylvén & Sundquist, 2012).

The positive relation between textual interpretation and textual output (see Weigle, 2004) suggests another important benefit of the *MYP Guide's* open-ended treatment of source texts. The power of a genre-based pedagogy in teaching learners how to interact with a variety of audiences and contexts can optimize the potential benefit that Weigle's findings suggest. One such pedagogy is central to the interpretation and use of text conventions, an assessment thread that appears explicitly in Assessment Criteria A, B, and D. Essentially, the more text types that learners analyze, the more empowered they will be to communicate effectively both within and outside the classroom in context (Hyland, 2007).

Comprehension and Interaction

As the Interaction Hypothesis (discussed in Yuksel & Inan, 2014) posits, interaction that facilitates negotiation of meaning facilitates language acquisition. Research on interactions (see Swain & Lapkin, 1998) examines synchronous forms of communication. Therefore, we can only draw clear connections between this line of research and Criterion C. Depending upon how an assessment task is contextualized, a connection with Criterion D might also emerge. However, Devitt (1997) outlines the many meanings of interaction, which "can refer to oral interaction between two or more people . . . or written interaction over the Internet. It can also refer to several kinds of interaction that take place in reading: the interaction between the reader and the text . . ." (p. 457).

Since the MYP *Guide* does not define the term, one can extend Devitt's argument to Criteria A and B (listening comprehension and reading comprehension). One could then argue that those criteria require negotiation of meaning, even though both engender unidirectional communication.

Another important consideration regarding how the four components of communicative competence may engage learners in interaction is to consider interactional competence (IC), a fifth competence that some language acquisition researchers support adding to Canale and Swain's framework. For example, Dings (2014) discusses alignment activity, a strategy that interlocutors develop when building their ICs. Simply put, alignment activity is any speech act that interlocutors engage in so that an interaction can be maintained. Undoubtedly, the MYP *Guide's* use of source texts to inspire output and interpersonal communication (most clearly related to Criterion C, though the Guide gives indication that interpersonal communication can also serve as evidence for Criterion D) increases the likelihood that learners will be able to engage in alignment activity and have meaningful interactions, because source texts provide some common experiences for the learners to draw upon when interacting with one another.

Output

As has been mentioned, learners are well positioned to interact in a variety of contexts if they experience a plethora of text types. Similarly, as Gass and Selinker (2008) attribute to Swain (1985, p. 249), "semantic processing leads to syntactic processing" and output as a process is central to learning (p. 328). Thus, the communicative tasks required to engage in each of the assessment criteria should be contextualized both as tasks to prove the occurrence of learning (i.e., summative) and tasks that cause learning to occur (Harklau, 2002), which can be described as formative. The MYP *Guide* proves alignment to this principle with its simultaneous explanation of summative and formative assessment tasks under "Subject-specific guidance" (pp. 42–47). The "Assessment" section of this literature review adds detail to this discussion.

Blending of Language Continuums in Language Processes: A Theoretical Perspective

Intercultural Competence

As has been discussed, the language continuums blend to form the language processes featured in the MYP *Guide*. Though the manner in which they blend can present some problems (see the following section on Assessment), they are noteworthy for their potential to promote intercultural competence, a model for which originally appeared in Byram (1997). Intercultural competence involves the attitudes, skills, and knowledge that someone holds in which high levels enable an individual to revisit personal values and beliefs when confronted with disparate beliefs without valuing one over the other. To develop this competence, individuals must develop skills of interpreting and relating, discovery and interaction, and critical awareness as outlined below:

- Skills of interpreting and relating (*savoir comprendre*): ability to interpret a document or an event from another culture, to explain it, and relate it to documents of one's own.
- Skills of discovery and interaction (*savoir apprendre/faire*): ability to acquire new knowledge of a culture and cultural practices and the ability to operate knowledge, attitudes and skills under the constraints of real-time communication and interaction.
- Critical cultural awareness (*savoir s'engager*): an ability to evaluate, critically and on the basis of explicit criteria, perspectives, practices and products in one's own and other cultures and countries. (Byram et al., 2001, pp. 6–17)

Most notably, learners are poised to develop intercultural competence when engaging in Assessment Criteria A and B. Within these criteria, Strand iii evaluates learners for their ability to engage with the source text by identifying ideas, opinions, and attitudes. This engagement requires *savoir comprendre* through the mere interpretation of text, *savoir apprendre/faire* through the construction of the personal response, and *savoir s'engager* through requiring that learners create personal responses. In this sense, the guide prepares learners to interact in a globalized world through the development of intercultural competence, an important consideration for MYP, which shares the IB mission that emphasizes the related construct of international-mindedness. Although intercultural competence is difficult to test empirically (Chun, 2011), the potential for learners to develop this competence can be furthered during language learning (i.e., under statements of inquiry, which are attached to each unit of study).

Contextualized Communication

Along these lines, another theoretical strength of blending language continuums to form language processes is that language acquisition is both socially mediated and cognitive (Atkinson, 2002). Each assessment criterion emphasizes context, and by extension, the social acceptability of texts both analyzed and produced. For clear examples of this consideration of context, please see Table A2's delineation of assessment criteria and sociolinguistic competence.

Multimodality and Multiliteracies

The last, and perhaps clearest, theoretical strength evident in how the language continuums are blended in the assessment criteria reflects the multimodality of digitally mediated communication, an increasingly evident form of input and platform for output that learners interact with in their daily lives. As Lotherington and Jenson (2011) comment, "literacy engages people in texts and discourses that traverse space and time on screens in which we can access and mix semiotic resources that include a multiplicity of languages" (p. 266). Thus, even though considering multiple inputs simultaneously can lead to cognitive overload for less proficient learners (Lotherington & Jenson, 2011; Mayer, Lee, & Peebles, 2014), learners must be prepared to interact with multimodal texts in the language to be acquired (i.e., L2) if for no other reason than the degree of proliferation of such inputs. In fact, learners interacting in multiple modes simultaneously is not limited to digitally mediated spheres. Face-to-face interactions involve visual and oral/aural modes (Lotherington & Jenson, 2011) and classroom interactions combine "literate, visual, and oral/aural modes of presentation" (Harklau, 2002, p. 336). Though practitioners should account for perceptions of text type when constructing multimodal assessments (see Chun, 2011), the MYP *Guide* justifiably supports the exploration of digital forms of communication.

Assessment

Sociocultural theory and pedagogy influence language acquisition classrooms (Mantero, 2002). Simply put, communicative tasks lead to the acquisition of language; therefore, communicative tasks should be used to both build and prove language knowledge. The multimodal and integrative nature of communication shrouds the practical execution of language assessments in complexity. Practical execution is the focus of the ensuing section.

The Blending of Language Continuums in Assessment Criteria: A Practical Perspective

Though the research discussed to this point largely supports how language continuums combine to yield multimodal assessment criteria/language processes, the selection of both the continuums and decisions regarding how they blend is not always intuitive in the *MYP Guide*. For example, it is unclear how the language continuums of listening and speaking, viewing and interpreting, reading comprehension, and writing materialized, despite the Guide's brief explanations. Furthermore, presentational speech is not explicitly accounted for, though the Guide alludes to some presentational tasks (see pp. 28–29); neither is synchronous interactive writing. Similarly, looping listening comprehension into speaking minimizes listening in comparison to reading. Because much research explores connections between reading and writing (e.g., Devitt, 1997; Fulcher, 1998; Hyland, 2007; Plakans, 2009; Weigle, 2004), it would be equally logical to combine reading and writing into a single continuum. Moreover, viewing and interpreting may be involved in the listening and speaking continuum and in the reading comprehension continuum depending on the context of a given language task. Consequently, it is unclear why viewing and interpreting have received their own language continuum.

The following sections outline the impact of the selection of these continuums on assessment objectives, highlighting the difficulties that they may present to learners and educators.

Authentic Assessment and the Limitations of Assessment Criterion A and Assessment Criterion C

The *MYP Guide* goes to great lengths to blend the areas of communication (oral, visual, and written) and the associated language continuums (listening and speaking, viewing and interpreting, reading comprehension, and writing) within the language processes as defined by the assessment criteria. Undoubtedly, the literature we discussed supports such a blend. However, to mandate multimodality in a given summative assessment disallows some assessment tasks that would be in alignment of Wiggins' (1989; 2011) "authentic assessment." As Frey, Schmitt, & Allen (2012) discuss, for an assessment to be authentic, the context must be realistic, and the task must be performance based and cognitively complex (p. 5). Although the constructs of all summative assessment criteria featured in the *MYP Guide* allow for the design of assessment tasks that are performance-based and cognitively complex, the current constructs of Criteria A and C do not allow for assessment tasks that encompass all realistic contexts. For example, Criterion A, responding to a text that is both spoken and visual (e.g., a commercial, a news clip, or a documentary), does account for auditory comprehension but does not allow for a summative assessment task that involves comprehension that is purely auditory. Given the ubiquity of situations that occur in real life in which people engage in such comprehension (e.g., listening to the radio, listening to announcements at airports or train stations, or even eavesdropping), mandating a multimodal source text is acceptable with respect to fostering and evaluating language acquisition but unnecessary. Similarly, it is acceptable but unnecessary that learners engage in interactions in response to spoken and/or written and/or visual texts as is mandated by Assessment Criterion C. Plenty of meaningful interpersonal interaction occurs between interlocutors in real life (e.g., meeting someone for the first time, asking for directions, or making weekend plans with a friend) that is not prompted by some sort of outside text. Even though the Guide allows learners to engage in an interaction that is not based on a source text in order to achieve Criterion D, the criterion does not explicitly require the evaluation of interactive skills, thereby supporting the aforementioned critique.

It is exceedingly important to note that these observations are not included in order to advocate the removal of the assessment threads that require learners to consider the conventions that are most appropriate to given text types. In fact, these considerations are essential to language acquisition. Hyland (2007) writes that “language is structured to achieve social purposes in particular contexts of use” (p. 148). The evaluation of the conventions of texts is central to understanding both the impact of how language is structured and how to structure it one’s self. The knowledge gained from these evaluations informs not only multimodal communication for learners, but also communication within a singular mode. Unimodal communication is realistic and authentic; although summative assessment tasks that exclude these types of communication do not ignore the principles of language acquisition or authentic assessment, there is no reason from a language acquisition perspective or an authentic assessment perspective that they be excluded.

Comprehensibility of Feedback for Learners: Areas of Communication in Combination

As Bigelow, Delmas, Hansen, and Tarone (2006) discuss, “Interaction gives learners opportunities to test hypotheses about the L2, receive negative and positive evidence, notice syntactic or lexical gaps in the L2, and in general negotiate both form and meaning . . .” (p. 669). In the classroom, this interaction does not simply occur when learners engage in interpersonal exchanges. Rather, every bit of feedback on any output provides learners with the opportunity to receive the necessary negative and positive evidence to improve their communication skills. Learners create work, and educators evaluate the work so that it is beneficial to inspiring change. However, whether feedback is understood ultimately determines its benefit. Proponents of the Noticing Hypothesis argue that attention to and noticing of errors are important to language acquisition (Mackey, 2006). Though this hypothesis needs further validation, it is logical that if learners receive some information intended as feedback, but the learner does not understand how that information relates to their objectives, the information would have no transformative power.

Though it is true that the areas of communication and language continuums blend in real life, the strands within Criteria A, B, and C are potentially confusing to learners. Each criterion involves receptive and productive skills, complicating learners’ comprehension of their own strengths or weaknesses with respect to individual macrolevel language skills when considering assessment results. For example, a learner engaged in a task to assess listening comprehension (Criterion A) might identify basic facts, messages, and main ideas and supporting details in a text (Strand i), but show no awareness for basic conventions (Strand ii), and have such limited language control that it is impossible to understand how well he or she engaged with the source text (Strand iii). Thus, if learners knows how well they perform in Criterion A, which involves visual comprehension, auditory comprehension, and some sort of language production, learners do not necessarily have enough information to understand how they engage in each of the macrolevel communicative skills (listening, speaking, writing). Furthermore, researchers do not understand fully the role that each macrolevel language skill plays when they are assessed in concert (Plakans, 2009).

Summative and Formative Assessment Link

Learners must have the opportunity to repair and modify output to produce uptake (Egi, 2010). This is a continual process that requires learners to be enrolled in a proficiency-based classroom, within which learners use communicative tasks as vehicles for language acquisition rather than as simple proof of acquisition (Mantero, 2002). The *MYP Guide* does advocate such an approach in the classroom by listing student evidence for each of the continuums (pp. 28–35) and by providing

suggested task types for formative and summative assessments (p. 43). However, it is unclear if the guide provides sufficient instruction to practitioners regarding how to structure the daily operations of the classroom. It is possible that a more detailed class outline in which sufficient formative work and peer work (see Pellet, 2012) were highlighted would help practitioners to internalize how to implement proficiency-based classrooms. Though consideration of the Approaches to Learning listed in the *MYP Guide* (pp. 40–41) does yield some additional insights into what learners need to experience in order to improve their language proficiencies, it is not clear that this is sufficient information for practitioners attempting to develop proficiency-based classrooms either. The relative inattention in the *MYP Guide* to the day-to-day progression within classrooms may yield situations in which practitioners use communicative tasks to prove learning rather than as they should be used: for language learning and acquisition (Mantero, 2002).

Conclusion

The current structure of the *MYP Guide* yields some confusion regarding its practical application. The language processes and language continuums specified in the *Guide* do not align directly with CEFR and ACTFL, two standard sets that are ubiquitous on an international level. Furthermore, the multimodality and integrative nature of assessment tasks present teachers with the rather arduous task of providing comprehensible feedback to learners. Meanwhile, little guidance is provided to teachers regarding how meaningful communicative tasks may be integrated into classrooms on a day-to-day basis. Still, the overall theoretical underpinnings of the *MYP Guide* are solid, directing teachers to provide learning opportunities for students in which they engage in meaningful communication, fostering the growth of their respective language proficiencies. Communication is shown accurately: as being complex, multimodal, and integrative.

APPENDIX C: WITHIN-DOCUMENT REVIEW TOOL

Sample screenshots from the online within-document review tool are presented below. If you would like a tour of the full review tool, please contact the lead author of this report at michael_thier@epiconline.org.

Global Proficiency Table: Six Phases (Page 25)

Directions: The review of the Six Phases articulated in the MYP Language Acquisition Global Proficiency Table, involves two areas of focus. **First**, please examine the **clarity** of the Six Phases and note whether descriptions in the Global Proficiency Table are *Clear* or *Not Clear*. Where the descriptions are *Not Clear*, please briefly describe in the textbox **either** the problem/challenge with the MYP Language Acquisition Global Proficiency Table or recommendations to make the descriptions clearer to enhance their usefulness to educators. **Next**, please examine the degree of **alignment** between the phase descriptions within the Global Proficiency Table and descriptions within the Assessment Criteria Rubric, and note where descriptions are *Aligned* (they are the same or similar enough) or where they are *Not Aligned* (they differ in substantive ways). Where the description are *Not Aligned* please note **either** the problem/challenge or recommendations to enhance the consistency of descriptions between the two documents. For this specific alignment comparison, please examine each phase described in the Global Proficiency Table holistically.

Clarity of the Six Phases in the Global Proficiency Table						
Are the 6 Phases CLEAR? (Clear = I know what evidence of this phase looks like)						
Clarity of Phases in the Global Proficiency Table						
	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6
Emergent						
Capable						
Proficient						
In instances marked 'Not Clear', briefly describe the problem/challenge OR provide a recommendation in the box below.						

Alignment Between the Global Proficiency Table Six Phases and the Assessment Criteria						
Considering the Criteria holistically, are the Phase-specific Language Acquisition Objectives ALIGNED with descriptions in the Assessment Criteria Rubrics at the 7-8 achievement level? (Aligned = the same or similar enough)						
Criteria	Alignment Between the Phase-Specific Language Acquisition Objectives and the Assessment Criteria Rubric at 7-8 achievement level					
	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6
A. Comprehension Spoken/Visual						
B. Comprehension Written/Visual						
C. Communicate in Response						
D. Use Language in Spoken or Written Form						
In instances marked 'Not Aligned', briefly describe the problem/challenge OR provide a recommendation in the box below.						

Within-Document Analysis Exit Questions

After completing your review of each progression within the MYP Language Acquisition Guide, please share your perspective on the collection of progressions as a whole by answering the questions below.

How effectively do the different progressions help you judge student placement and plan for instruction across the middle years?

APPENDIX D: CROSS-DOCUMENT REVIEW TOOL

Sample screenshots from the online cross-document review tool are presented below. If you would like a tour of the full review tool, please contact the lead author of this report at michael_thier@epiconline.org.

MYP Language Acquisition Guide

Directions: For this review, you will examine the degree to which two frameworks conceptualize and describe a progression of learning in relation to research on language acquisition and language learning. This first tab provides a space for you to examine the **MYP Language Acquisition Guide** holistically and note the degree to which the MYP Language Acquisition Guide reflects aspects of language learning identified below. For each category of language learning, such as Interpersonal Communication, please use the drop-down options to note the extent to which an item, such as "oral communication", is addressed in the MYP Language Acquisition Guide by selecting either: "Not at All", "Too Little", "Enough", or "Too Much". You will use the same process to examine the Common European of Reference for Languages in the second tab. After examining both the MYP Language Acquisition Guide and the Common European Framework of Reference for Languages, please respond to the prompts in the EXIT QUESTIONS tab before submitting this document to EPIC.

Interpersonal Communication

To what extent are these aspects of Interpersonal Communication addressed in the MYP Language Acquisition Guide?	
Interpersonal oral communication	
Interpersonal written communication (e.g., letters, emails, texts, online communities)	

Defining Phases of Language Acquisition

To what extent are the following addressed to define phases of language acquisition in the MYP Language Acquisition Guide?	
Language complexity increases as learners move from Phase 1 forward	
Topics and content also become more broad and varied	
Communication acts, not syntactic and lexical accuracy, are the focus	
Reflects the contemporary proficiency movement in language acquisition that rejects previous approaches (e.g., grammar-translation)	
Progression is logical and qualified by expertise	

Four Communicative Competencies

To what extent are the four communicative competencies addressed in the MYP Language Acquisition Guide?	
Grammatical: A person's ability to use language appropriately and according to convention	
Sociolinguistic: A person's ability to use language appropriately in a given communication context	
Discourse: A person's ability to produce cohesive and complex thoughts	
Strategic: A person's ability to use strategies to compensate for gaps in knowledge and misunderstanding	

Cross-Document Analysis Exit Questions

Directions: After completing your holistic review of both the MYP Language Acquisition Guide and the Common European Framework of Reference for Languages, please share your perspective on strengths unique to each language acquisition framework by responding to each of the two prompts below. Thanks again for committing your time and effort for this review! Your insights are valued and will help inform potential improvements to the MYP Language Acquisition Guide.

Are there unique strengths within the **MYP Language Acquisition Guide's** progressions of learning compared to the other framework of language learning? If so, please describe these unique characteristics below.

--	--	--	--	--	--	--	--

Are there unique strength within the **Common European Framework of Reference for Languages** that, if incorporated into the MYP Language Acquisition Guide, would enhance its usefulness for educators? If so, please describe these strengths and provide any suggestions on how they could be used to improve the MYP Language Acquisition Guide.

--	--	--	--	--	--	--	--

APPENDIX E: CONTINUUM-SCHOOL TEACHER INTERVIEW PROTOCOL

Thank you for taking time to discuss your experiences as an IB educator with me today. Before we begin, I wanted to check if you reviewed the document that EPIC sent with findings from an analysis of language-acquisition expectation alignment across the PYP, MYP, and DP. Did you have a chance to review that document?

If yes: Terrific! Thank you very much. Do you have any questions before we begin the interview?

1. Can you briefly share any experiences that you, your students, or colleagues have noted as being easy or difficult in their attempts to understand how the programmes' expectations differ from each other's?

2. How consistent to your experience as an IB educator is the document's description of alignment across the programmes in the following four areas?

- conceptualizing effective language teaching and learning
- expectations of meaningful communication in authentic contexts
- emphasizing intercultural competence
- assessment approaches that involve both processes (formative) and products (summative) of learning

3. What effect on your role as an IB educator, if any, would there be if IB made its language-acquisition terminology consistent across the PYP, MYP, and DP? For example, what if IB implemented the same assessment criteria or communicative competencies across the three programmes?

4. Other than what we've discussed so far, what specific areas of the document, if any, *conformed to your experience* as an IB educator?

5. Other than what we've discussed so far, what specific areas of the document, if any, *did not conform to your experience* as an IB educator?

6. Is there any additional information you'd like to provide about the progression document?

Thank you very much for participating in this research project.

APPENDIX F: HEAD OF MYP DEVELOPMENT INTERVIEW PROTOCOL

Thank you for taking the time to speak with me today. In this final phase of our research project with IB on *Alignment and Coherence of Language Acquisition Development in the Middle Years Programme*, we want to examine two aspects of this research. First, we want to hear about any recommendations you might have to improve or otherwise change the MYP Language Acquisition *Guide*. Second, we want to probe further in a few areas where our data from previous phases detected discrepancies or inconsistencies. Importantly, I want to remind you that the review has also uncovered a number of strengths such as:

- interpersonal communication
- defining phases of language acquisition
- meaningful communication in a variety of authentic contexts
- intercultural competence

Though our final report will explore both strengths and weaknesses, this phase of the project emphasizes your opinions on any changes or refinements that the MYP *Guide* should undergo to promote good teaching practice and efficiency in placing students and evaluating their progress in language acquisition.

Do you have any questions before we begin?

Great.

1. Throughout this project, we heard reviewers (MYP educators and IB staff) discuss a *need for more explicit definition or exemplification* in the *Guide*. How, if at all, should IB address or provide more explicit definition and/or exemplification to the *Guide* and/or related MYP documents to ultimately aid implementation?

Thank you. Now we will move onto our next question.

2. Many of our reviewers noted instances in which the *Guide* or related MYP documents *did not demonstrate shared language*. Reviewers identified inconsistent application of terminology or inadvertent conflating of potentially overlapping terms (e.g., language continuums and language competencies, strands, progressions, etc.). What, if any, recommendation would you have to address this matter?

Thank you. Now we will move onto our next question.

3. Reviewers were somewhat divided on the *extent to which differences between the three main IB programmes (i.e., PYP, MYP, DP) raised concerns* for students, teachers/administrators, students' families, and other stakeholders. For example, the crucial MYP term "communicative processes" does not exist for language acquisition documents in the PYP or DP. Moreover, the assessment structures for the three programmes vary considerably from little being formalized in the PYP to a highly specified DP assessment. Some participants in our research felt that such variation can thwart some schools' attempts to place students properly. Meanwhile, some reviewers recognized

the need for a different approach in the DP for “grown-up kids.” Given a range of viewpoints, to what extent should IB’s programmes align with one another in terms of shared language, shared assessment practices/criteria, and other ways?

As a follow-up, how can IB help practitioners and schools achieve that articulated aim?

Thank you. Now we will move onto our next question.

4. Our data reflect multiple reasons that lead stakeholders to *ponder revisions to assessment*, although such considerations did not occupy a very large proportion of our data. Examples included:

- merging assessments so language learners could engage in both input and output on a single assessment to avoid complicating feedback loop
- considering the appropriateness of using mother tongue and/or language of instruction during summative assessments
- reexamining word-length ranges
- the tension between newer assessment modes, such as those in digitally mediated environments or synchronous interactive writing, and older modes such as attending to errors in vocabulary, grammar, and syntax

Given these reasons, how should IB address some or all of these issues that pertain to revising assessment?

Thank you.

5. The next question stems from what might be a distortion in the data or the insights of one practitioner adding value to the study. In light of a clearly recognized strength across data collection methods that pointed to the *Guide* emphasizing communication in authentic contexts, one participant voiced concern about whether IB teachers truly have the time to “immerse their students in language learning every single day.” Given this discrepancy between data sources in our study, to what extent does the IB believe that its practitioners can sufficiently immerse students daily in authentic language learning contexts?

6. Is there anything else that you would like to tell us regarding changes or refinements that would be beneficial for users of the *Guide*?

Thank you for your time.