Technical Manual

for the

Contaminated Sites Approved Professionals (CSAP) Society

Numerical Standards and Risk-based Standards Oral Assessments©

January 2024

Table of Contents

Preface

Purpose of the Manual

The purpose of the Technical Manual for the CSAP Oral Assessments is to explain the purpose of the CSAP Oral Assessments, how these assessments were developed and what the assessment results are used for. Throughout the design and development process, the CSAP Society is committed to ensuring a valid, reliable, and fair assessment process that is based on rigorous psychometric testing standards as recommended by the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014).

Purpose of the CSAP Oral Assessments

In line with the mission of the Society of Contaminated Sites Approved Professionals (CSAP) of British Columbia "to serve government, the public and industry by evaluating and advancing the practice and quality of contaminated sites management within BC's regulatory framework" (CSAP, n.d.), the CSAP Society designed and developed two oral assessments to be used to measure application of technical knowledge and decision-making in the areas of numerical standards and risk-based standards, respectively. In the September 2023 administration, these oral assessments were administered for the first time in place of the existing knowledge-based technical exams for numerical standards and risk-based standards, which were used to measure technical knowledge. To become an Approved Professional (AP), candidates are required to pass the knowledge-based regulatory exam and either or both the numerical-based and risk-based standards oral assessments, based on the designation for which they are eligible.

Oral assessments are performance-based assessments (Swanson, Norman, & Linn, 1995) or judge-mediated examinations (Stone, Beltyukova, & Fox, 2008) that have been successfully used to assess application of knowledge and decision making in health care professions for over a century (Tekian & Yudkowsky, 2009). The decision to replace the knowledge-based written exams with performance-based oral assessments was informed by a comprehensive practice analysis which is detailed in the practice analysis section of this report. In general, the practice analysis provided explicit evidence of the alignment of the primary tasks of an AP with the most appropriate method of assessment.

The purpose of the CSAP Oral Assessments is to determine whether an examinee has sufficient technical knowledge in the assessment of contaminated sites, and the ability to make sound technical interpretations and decisions within the Contaminated Sites Regulation framework to make applications as either a Numerical Standards Approved Professional or a Risk-based Standards Approved Professional. Eligible applications are defined under the BC Ministry of Environment and Climate Change Strategy (ENV) Protocol 6.

The purpose of the CSAP Oral Assessments was developed by a team of numerical standards APs and risk-based standards APs, under the guidance of the psychometric consultant contracted by the CSAP Society. The APs that were involved with the development of the CSAP Oral Assessments have between 20-39 years of experience in the respective CSAP fields. The psychometric consultant, who wrote this technical manual, has over 11 years of experience in providing his expertise in developing and evaluating high-stakes certification assessments. He has a PhD in Measurement, Evaluation, and Research Methodology from the University of British Columbia. The role of the psychometric consultant is to ensure high standards of assessment development are met through evidence-based processes that ensure assessments are reliable, valid and fair based on rigorous psychometric testing standards (AERA, APA, & NCME, 2014).

Design and Development of the CSAP Oral Assessments

Introduction

The design and development process of the CSAP Oral Assessments adheres to strict guidelines detailed in the Standards for Educational and Psychological Testing (AERA, APA, NCME, 2014):

- Clearly define the purpose of the assessment and the claims one wants to make about the test takers.
- Develop and conduct practice analysis to confirm domains of knowledge to be tested.
- Develop test specifications or assessment blueprints consistent with the purpose of the assessment and the domains of knowledge supported by the practice analysis.
- Develop specifications for question formats and numbers of questions needed to adequately sample the domains of knowledge supported by the practice analysis survey.
- Develop assessment questions that provide evidence of the measurable-behavior indicators detailed in the assessment blueprint.
- Review assessment questions and assembled assessment forms so that each question has a single best defensible answer and assesses content that is relevant to the job.
- Review assessment questions and assembled forms for potential fairness or bias concerns, revising or replacing items as needed to meet standards.

Assessments developed to inform certification decisions, as is the case for the CSAP Oral Assessments, are intended to convey the extent to which the test taker (candidate for the credential) has a sufficient level of knowledge and/or skills to perform important occupational activities in a safe and effective manner (AERA, APA, & NCME, 2014).

“Testing used in the processes of licensure and certification, which is generally referred to as credentialing, focuses on a candidate’s current skill or competence in a specified domain. In many occupations, individual practitioners must be licensed by governmental agencies. In other occupations, it is professional societies, employers, or other organizations that assume responsibility for credentialing” (AERA, APA, & NCME, 2014; p.169). The latter case is pertinent to the CSAP Society.

Certification assessments are “intended to provide the public, including employers and government agencies, with a dependable mechanism for identifying practitioners who have met particular standards. The standards may be strict, but not so stringent as to unduly restrain the right of qualified individuals to offer their services to the public. Certification testing also serves to protect the public by excluding persons who are deemed to be not qualified to do the work of the profession or occupation” (AERA, APA, & NCME, 2014; p.175).

Certification assessments, like the CSAP Oral Assessments, often included in the larger certification process— “typically includes educational and experiential requirements—because it represents a standardized, uniform opportunity to determine if a test taker has acquired and can demonstrate adequate command of a domain of knowledge and/or skills that the

profession has defined as being important or necessary to be considered qualified to enter the profession" (ETS, 2023; p. 11).

The main source of validity evidence for certification assessments comes from the alignment between what the profession defines as essential or critical knowledge, skills, and/or competencies important for safe and effective practice and the content of the assessment (AERA, APA, & NCME, 2014). This alignment is conducted through expert judgement by practitioners and other stakeholders in the profession who have an informed perspective about key knowledge, skills, and/or competencies needed to become certified to practice in the field.

The knowledge, skills, and/or competencies that the assessment requires the test taker to demonstrate are typically justified by a comprehensive practice analysis as being necessary for safe and effective practice (AERA, APA, & NCME, 2014).

Framework for the design of the oral assessments

The design of the oral assessments is guided by a comprehensive assessment design framework called Evidence Centered Design (ECD), which was introduced by Mislevy, Steinberg and Almond (2003). This framework has been successfully used to design assessments that measure complex knowledge and skills and provides a structured process for collecting validity evidence throughout the design process (Riconscente, Mislevy, & Corrigan, 2015). ECD is based on three key questions: 1) What knowledge, skills, or abilities should be assessed? 2) What behaviors or performances should reveal those constructs? 3) What tasks or situations should elicit those behaviors? In other words, ECD provides "a strong foundation for the validity argument by requiring documented, explicit linkages among the purpose for a test, the claims made about test takers, the evidence supporting those claims and the test takers' responses to tasks that provide the evidence" (Zieky, 2014; p. 80). Under the ECD framework, assessment is the "process of reasoning from the particular things people make, say or do to draw inferences about their knowledge, skills and abilities" (Zieky, 2014; p. 40). Therefore, the assessment design process requires a chain of reasoning that is effectively organized into 5 groups of activities or layers: 1) domain analysis, 2) domain modeling, 3) conceptual assessment framework, 4) assessment implementation, and 5) assessment delivery.

In the domain analysis layer, the purpose of assessment is clearly defined and then a practice analysis is conducted to comprehensively define the domain of practice operationalized through a list of key tasks performed by the numerical standards AP and the risk-based standards AP. This transitions into the domain modeling layer which requires the generation of explicit claims about what the AP is expected to know and be able to do. Next, the construction of a valid assessment blueprint is conducted which provides clear direction on what will be assessed and how it will be assessed. This occurs under the conceptual assessment framework layer in which the following questions are considered: 1) What is being measured, 2) How is it being measured? 3) Where is it being measured? 4) How much of it is needed to measure it?

In the assessment implementation layer, evidence required to support claims made about what the candidate is expected to know and be able to do is contemplated. Specifically, tasks or items are developed, along with scoring rules, to provide the desired evidence within the constraints of the assessment program. Finally, in the assessment delivery layer, the administration and scoring process is clearly established to ensure that the assessment process allows the accumulation of evidence for validity, reliability, and fairness.

Assessment Development Process

Practice Analysis

According to psychometric experts, practice analysis is conducted to inform the following key assessment development activities: 1) delineating the primary tasks performed in the profession, 2) defining the knowledge and skills required for successful performance, 3) selecting appropriate methods for assessing the tasks and/or the knowledge and skills that comprise this domain, 4) designing practice-related assessment items or tasks, and 5) defining an appropriate scoring methodology that is aligned with the purpose of the assessment (Clauser & Raymond, 2017). The practice analysis process allows the collection of strong validity evidence for the purpose of the assessment, the alignment between the tasks performed in practice and what is assessed, as well as the alignment between what is assessed with how it is assessed (the method of assessment).

The CSAP practice analysis process began with a 2-hour training session that was delivered virtually by the psychometric consultant to two subject matter experts who were the team leaders of their respective CSAP specialty (a risk-based AP and a numerical standards AP). The SMEs, who are APs, have between 20 and 39 of years of experience working as APs.

In the training session the psychometric consultant provided some background information on what practice analysis is, why it is needed and what it entails. During the training the psychometric consultant engaged in a discussion with the team leaders on the nature of the work performed by the APs as well as the context in which the work is performed. At the end of the session, the SMEs were tasked with developing a statement of purpose for their CSAP specialty (i.e. numerical standards and risk-based standards). The statement of purpose is the recommended first step in conducting a practice analysis as it provides clear direction on what is being assessed and why. These statements allow the investigation of whether the oral assessment method is the most appropriate way to assess the construct of interest. The statements of purpose, which were initially generated independently by the two SMEs, was reviewed by the psychometric consultant at the next meeting and was revised collaboratively by the SMEs to ensure consistency and clarity of purpose. The psychometric consultant also provided feedback to the team leaders to ensure a clear and strong alignment between the construct to be assessed (i.e. application of technical knowledge and decision making) and the best way to assess it (oral assessment). The statements of purpose are summarized in Table 1 below.

Table 1: *Statements of Purpose for Numerical Standards and Risk-based Standards*

	Statement of Purpose
Numerical standards	The purpose of the numerical standards oral assessment is to determine whether an examinee has sufficient technical knowledge in the assessment of contaminated sites, and the ability to make sound technical interpretations and decisions within the Contaminated Sites Regulation framework to make applications as a Numerical Standards Approved Professional. Eligible applications are defined under the BC Ministry of Environment and Climate Change Strategy (ENV) Protocol 6.
Risk-based standards	The purpose of the risk assessment oral assessment is to determine whether an examinee has sufficient technical knowledge in the assessment of contaminated sites, and the ability to make sound technical interpretations and decisions within the Contaminated Sites Regulation framework to conduct work as a Risk Assessment Standards Approved Professional including the submission of applications to the BC Ministry of Environment and Climate Change Strategy (ENV). Eligible work is defined under the ENV Protocol 6.

To ensure alignment between the construct to be assessed and how best to assess it, the psychometric consultant conducted a thorough research literature search on oral assessment use in licensure and certification examination contexts. Through this review it was established that the purpose of the CSAP oral assessment is well aligned with the documented purpose of oral assessments as stated in the literature, which is to "explore an examinee's thinking in order to assess skills such as critical reasoning, problem solving, judgment, and synthesize material" (Tekian & Yudkowsky, 2009; p. 269).

Next, the SMEs were instructed to generate a list of tasks that numerical standards APs and risk-based standards APs perform when reviewing applications. To assist with this work the SMEs were encouraged to consult existing documents and job descriptions, as well draw upon their own experience working with newly certified APs. The psychometric consultant also asked them to obtain daily or weekly log of activities from newly certified APs. After several discussions with the SMEs it was determined that the AP tasks that were generated for the written exams have been based on authorized activities defined and mandated by the British Columbia Ministry of Environment and Climate Change Strategy (ENV). The tasks an AP engages in depends on the applications they receive for a contaminated sites service. The primary tasks that have been generated based on the authorized activities are comprehensive and cover all possible tasks that a numerical standards and risk-based standards AP may perform. As a result of this information and after reviewing the primary tasks, the SMEs recommended using the primary tasks as the final task list for risk-based and numerical standards, respectively.

The next step in practice analysis was to evaluate the tasks. Typically a frequency and importance rating scale is used to determine which tasks are conducted more frequently than other tasks as well as which tasks are considered more important than other tasks (APA et al., 2014). However, after several discussions with the SMEs it was determined that AP tasks cannot be rated based on frequency or importance due to the nature of the work. Specifically, the tasks that APs engage in are dependent on the applications they receive, and that no tasks are more frequently performed than others. Also, all tasks are important to perform so it does not make sense to rate tasks by importance. In addition, the SMEs stated that all tasks must be performed appropriately as there are negative consequences for failing to perform the tasks. Random or non-random audits are conducted to check that APs are conducting their work appropriately. Failure to perform the tasks appropriately results in disciplinary action. In light of the absence of frequency and importance ratings the psychometric consultant asked the SMEs to provide a weighting percentage for each primary task category instead. The percentages would correspond to the proportion of items or questions that would be included in a given assessment. The higher the percentage for a given task category, the greater its importance in the assessment, and consequently, the more items or questions will be included.

Next the SMEs were instructed to develop a list of knowledge statements. To help generate knowledge statements the SMEs reviewed an existing document that provided a list of knowledge statements for the minimally qualified AP candidate for numerical standards and risk-based standards, respectively. After reviewing these documents the SMEs approved a final version of a general knowledge statement as well as list of specific knowledge statements. The general knowledge statements for numerical standards and risk-based standards are provided in Table 2 below.

Table 2: *General Knowledge Statements for Numerical standards and Risk-based standards*

	General Knowledge Statement
Numerical standards	A successful candidate needs to understand and have the ability to apply combined aspects of soil science, environmental engineering, hydrogeology, environmental chemistry and basic risk assessment principles to review standards assessments and applicable documents to make recommendations based on application of numeric-based standards of the BC CSR. Candidates are also expected to have a general understanding of related areas.
Risk-based standards	A successful candidate needs to understand and have the ability to apply combined aspects of ecology, toxicology and environmental chemistry to review of human health and ecological risk assessments and applicable documents to make recommendations based on application of risk-based standards of the BC CSR. Candidates are also expected to have a general understanding of related areas such as for example, basic contaminant transport in various media.

Once the final task and knowledge statements were approved, the next step of the practice analysis was to develop a survey to administer to the CSAP field for validation purposes. The surveys for numerical standards and risk-based standards are included in Appendix A and B, respectively. Both surveys were administered to risk-based and numerical standards APs by the CSAP Society.

The practice analysis survey for numerical standards was completed by five participants (response rate of 100%). Sixty percent of participants identified as numerical standards APs, while the remaining identified as both risk-based and numerical standards APs. Twenty percent of the participants reported a female gender identity, while the rest reported a male gender identity. Eighty-three percent reported English as their first language.

The practice analysis survey for risk-based standards was completed by six participants (response rate of 100%). Sixty-seven percent of participants identified as risk-based APs, while the rest identified as both risk-based and numerical standards APs. Thirty-three percent of the participants reported a female gender identity while the rest reported a male gender identity. Eighty-three percent of the participants reported English as their first language.

The results of both surveys are summarized in Appendix C and D, for numerical standards and risk-based standards, respectively.

Assessment Blueprint Development

After administration of the practice analysis survey the summarized results were presented to the SMEs for review. Based on feedback from the survey respondents, the statements of purpose and general knowledge statement was revised for both numerical standards and risk-based standards. The changes are summarized in Tables 3 and 4 below.

Table 3: *Revised Statements of Purpose for Numerical standards and Risk-based standards*

	Revised Statement of Purpose
Numerical standards	The purpose of the numerical standards oral assessment is to determine whether an examinee has sufficient technical knowledge in the assessment and remediation of contaminated sites, and the ability to make sound technical interpretations, judgements, and decisions within the Contaminated Sites Regulation framework to make applications as a Numerical Standards Approved Professional. Eligible applications are defined under the BC Ministry of Environment and Climate Change Strategy (ENV) Protocol 6, and the tasks specific to a Numerical Standards Approved Professional are defined in the Practice Analysis below.

Risk-based standards	The purpose of the risk assessment oral assessment is to determine whether an examinee has sufficient technical knowledge in the risk assessment of contaminated sites, and the ability to make sound technical interpretations, judgements, and decisions within the Contaminated Sites Regulation framework to conduct work as a Risk-based Standards Approved Professional including the submission of applications to the BC Ministry of Environment and Climate Change Strategy (ENV). Eligible work is defined under the ENV Protocol 6.

Table 4: *Revised General Knowledge Statements for Numerical standards and Risk-based standards*

	Revised General Knowledge Statement
Numerical standards	A successful candidate needs to understand and have the ability to apply combined aspects of soil science, geology, environmental engineering, hydrology, hydrogeology, environmental chemistry, biochemistry, contaminant fate and transport and basic risk assessment principles to review standards assessments and applicable documents to make recommendations based on application of numeric-based standards of the BC CSR. Candidates are also expected to have a general understanding of related areas such as sampling and laboratory testing methods, statistical analysis, contaminants associated with specific site activities and methods to remediation them.
Risk-based standards	A successful candidate needs to understand and have the ability to apply combined aspects of ecology, toxicology, environmental chemistry and basic contaminant transport in various media to review human health and ecological risk assessments and applicable documents (e.g., performance verification plans) to make recommendations based on application of Risk-based standards of the BC CSR. Candidates are also expected to have a general understanding of related areas such as for example, site investigation and characterization (e.g., sampling) for various media and data gaps that will limit the risk assessment and create uncertainty for the risk controls.

The blueprint or test specifications is a detailed written plan that provides all the information needed to develop the oral assessment, from the content that will be assessed and how much each content area will be weighted, to how the content will be assessed, administered, scored, and reported (Parshall & Brunner, 2017). In the measurement research literature, two types of test specifications are described: process-based and content-based (AERA, APA, & NCME, 2014). The focus of process-based test specifications is on defining the tasks, behaviors and activities required for safe and effective practice, whereas the focus of content-based test specifications is on defining the knowledge, skills, and abilities required for safe and effective

practice. The latter lends itself to the development of knowledge-based assessments in which knowledge and skills are measured, rather than the ability to perform a task, whereas the former lends itself to the development of performance-based assessments in which the behavior of candidates can be observed and/or they can demonstrate their ability to perform a given task.

According to the Standards for Educational and Psychological Testing (AERA et al., 2014), an effective blueprint must include detailed information on five specifications: content, format, psychometric, scoring, and test administration. Therefore the blueprint for the CSAP oral assessments was developed according to these guidelines.

The psychometric consultant began the blueprint development process with a 2-hour orientation on how to develop a blueprint. After a Q&A session the SMEs were then instructed to provide their input on the five specifications. Both SMEs agreed that the process-based test specification is most appropriate for the blueprint development of the oral assessments since the focus is on application of technical knowledge and ability to make appropriate decisions. The following input was provided by the SMEs in regards to this matter: “The assessment will be process-based – assessing ability to apply knowledge; the candidate will need to know technical knowledge to apply it, but the evaluation will be open-book with a reference list provided in advance that the candidate can bring into the assessment with them”.

Content Specifications

The first step in developing the blueprint is to clearly describe the content specifications, which delineate the purpose of the assessment and the scope of the content or domain that will be measured (AERA et al., 2014). Information on content specifications was provided by the practice analysis work as well as the results of the practice analysis survey. Specifically, the tasks that were described in the practice analysis and which were validated in the survey informed the content of the oral assessments. Based on the input provided by the SMEs as well as feedback from the survey participants, it was determined that the “Primary Tasks” will be measured in the oral assessments. The weights for each primary task category was informed by survey respondents. The final version of the content specifications are provided in Tables 5 and 6.

Table 5: *Content Specifications for Numerical standards*

Primary Task Category	Specific Tasks	Weight
Historical and Visual Site Information	a) APEC and PCOC: Identify all applicable potential APEC and PCOC based on review of existing information from various sources and based on assessment of site conditions observed during a site reconnaissance.	10%
Assessment of Affected Media and Migration Pathway	a) Soil: Interpret site geology, soil stratigraphy and depositional environment. b) Hydrogeology: Assess groundwater flow and contaminant transport (dissolved and non-aqueous phase liquids – NAPL). c) Surface hydrology: Interpret significance of precipitation on a contaminated site in terms of contaminant transport (surface water, groundwater, soil and sediment). d) Sediment: Interpret sediment characteristics and its significance for contaminant distribution and release. e) Soil vapour: Understand soil vapour concentrations and migration. f) Air: Understand impact on indoor and outdoor air quality by dust and vapours from site contamination. g) Biota: Understand the significance of observations such as stressed vegetation and effects on aquatic life.	19%
Contaminant Characteristics	a) Chemistry and biochemistry: Interpret physical, chemical and biological properties of contaminants and their significance on fate, transport, treatment and relative human health and ecological risks. b) Chemical composition of mixtures: Understand the significance of chemical compositions of common types of contamination substances including but not limited to: fuels, lubricants, solvents, paints, wood preservatives, coal tar, metal plating, and landfill leachate. c) Sources of Contamination: Be familiar with common residential, commercial and	15%

	industrial activities that may result in site contamination including but not limited to: Fuel storage and handling, metal fabrication, wood preservation, solvent cleaning, coal gasification, and landfilling.	
Investigation Approach and Methods	a) Sampling rationale: Interpret available information to develop a defensible sampling rationale that will satisfy the investigation objectives. b) Sampling plans: Assess sampling plans to determine whether they are consistent with the investigation objectives and sampling rationale. c) Sampling techniques: Understand the significance of the use of proper equipment and methods for sampling of soil, sediment, groundwater, surface water and soil vapour. d) Field observations and records: Assess field records in terms of adequacy for data interpretation included but not limited to: Borehole logs, well installation details, visual/olfactory signs of contamination, sampling details, etc. e) Laboratory testing methods: Understand applicability and limitations of common laboratory sampling methods including but not limited to: Gas chromatography, gas chromatography/mass spectroscopy, infrared spectroscopy, petroleum analytical methods (e.g., LEPH/HEPH vs. EPH). f) Field screening techniques: Understand applicability and limitations of common laboratory sampling methods including but not limited to: soil vapour headspace, immunoassay, colorimetric, pH/conductivity/temperature, X-ray fluorescence. g) QA/QC practices: Assess field and laboratory work in terms of acceptable QA/QC methods and interpretation.	15%
Data Synthesis and Interpretation	a) Data integration and presentation: Assess the investigation data in terms of adequate presentation in tables and figures.	15%

	b) Adequacy of testing: Review sampling programs to assess the adequacy of the testing performed (number, type and location of samples). c) Nature and extent of contamination: Assess APEC and AEC: number, types, characteristics, PCOC, delineation. d) Nature and extent of migration pathways: Assess migration pathways: types, characteristics, preferential routes, relative importance. e) Background conditions: Assess regional and local background conditions.	
Risk Assessment Principles and Screening	a) Problem formulation: Identify/screen sources, exposure pathways, receptors b) Acceptable risk : Carcinogens Non-carcinogens. c) Exposure scenarios: Interpret current and future site uses. d) High risk: Recognize imminent and high risk to human health and environment, and immediate risks to public welfare (e.g., explosion hazard, etc.). e) Screening Level Risk Assessment: concentration trend analysis, exposure scenarios, soil leachate assessment, groundwater transport assessment.	7%
Remediation Design, Implementation and Confirmation	a) Remediation techniques: Be familiar with common soil, sediment, groundwater, water and soil vapour remediation methods. b) Remedial design: Understand technical, regulatory and cost aspects of common remediation methods, and be able to evaluate the selection of appropriate alternatives. c) Remediation implementation: Understand health and safety standards, construction techniques/constraints, monitoring requirements, and requirements associated with off-site transport and disposal of contamination and record keeping. d) Remediation Confirmation: Assess confirmatory sampling program and results	15%

	in terms of adequacy and trend analysis to demonstrate the site meets the applicable requirements of a remediated site.	

Table 6: *Content Specifications for Risk-based standards*

Type of Assessment	Primary Task Category	Specific Tasks	Weight
Ecological Risk Assessment	Problem Formulation	a) Risk Assessment Planning b) Integration of Available Information c) Identification of stressors d) Potentially Exposed Receptors a. Complete and incomplete pathways b. Risk controls e) Selecting Assessment and Measurement Endpoints f) Conceptual Models g) Data Gap Analysis h) Sampling and Analysis Plan	15%
	Exposure Assessment	a) Characterization of Exposure b) Evaluating Data and Models for Analysis a. Strengths and Limitations of Different Types of Data b. Literature Data – relevant species, study conditions c. Site Data/Observations - measurement and assessment endpoints; species diversity, richness, abundance c) Measurement and/or Modeling Studies	12.5%
	Effects Assessment	a) Quantitative and Qualitative Site Observations a. Terrestrial Receptors b. Aquatic Receptors b) Bioassays a. Field studies b. Laboratory toxicity tests c) Toxicity Reference Values a. Selection	12.5%

		b. Derivation d) Ecosystem – context of scale relative to contaminated sites e) Ecological Responses a. Stressor-Response Analysis b. Establishing Cause-and-Effect Relationships c. Linking Measures of Effect to Assessment Endpoints	
	Risk Characterization	a) Quotient Method b) Observation Method c) Weight of Evidence d) Reporting Risks	10 %
	Uncertainty Analysis	a) Identifying Major Types of Uncertainty b) Use of Uncertainty Factors c) Sensitivity Analysis	
	Risk Management and Requirements	a) Performance Verification Plans b) Risk Management c) Other Risk-based Submission Requirements	
Human Health Risk Assessment	Problem Formulation	a) Data Collection a. Background Information Useful for Data Collection b. Review of Available Site Information c. Addressing Modeling Parameter Needs d. Preliminary Identification of Potential Human Exposure e. Strategy for Sample Collection f. QA/QC Measures b) Data Evaluation a. Combining Data Available from Site Investigations c) Evaluation of Analytical Methods d) Evaluation of Quantitation Limits e) Chemicals of Potential Concern a. Comparison of Samples with Criteria/Guidelines b. Comparison of Samples with Standards	15%

		f) Potentially Exposed Receptors g) Potential Exposure Pathways a. Complete and incomplete pathways b. Risk controls h) Conceptual Model i) Data Gap Analysis	
	Exposure Assessment	a) Characterization of Exposure Setting a. Characterize Physical Setting b. Characterize Exposed Receptors c. Identification of Exposure Routes d. Identification of Reasonable Maximum Exposure b) Quantification of Exposure: Determining Exposure Concentrations a. Estimation of Chemical Intakes b. Exposure Concentrations in Various Media c. Combining Chemical Intakes Across Pathways	15%
	Toxicity Assessment	a) Types of Toxicological Information Considered in Toxicity Assessment b) Toxicity Assessment for Noncarcinogenic Effects c) Toxicity Assessment for Carcinogenic Effects d) Identifying Appropriate Toxicity Values for Site Risk Assessment e) Evaluating Chemicals for which no Regulatory Toxicity Values are available	10%
	Risk Characterization	a) Quantifying Risks a. Risks for Individual Substances b. Risks for Multiple Substances b) Combining Risks Across Exposure Pathways c) Consideration of Site-Specific Human Studies d) Risk Characterization Results e) Risk controls	10%
	Uncertainty Analysis	a) Identifying Major Types of Uncertainty b) Use of Uncertainty Factors	

		c) Sensitivity Analysis	
	Risk Management and Requirements	a) Performance Verification Plans b) Risk Management c) Other Risk-based Submission Requirements	

Regarding the scope of the content, the SMEs stated the following: "…the content is not intended to measure the administrative requirements of being an AP, e.g., maintaining parent organization membership, PD hours, nor to assess regulatory process, which is covered under the current CSAP regulatory exam". Therefore, as indicated in the statement of purpose, the scope of the content for both oral assessments are limited to the application of technical knowledge and decision-making.

An important decision to make when designing content specifications is related to whether to directly assess the tasks or to assess the knowledge, skills, and abilities required to perform the tasks. The SMEs agreed that it is necessary to directly assess how the candidates perform the tasks. This will help inform the format specifications.

Format Specifications

The focus in specifying the format of the oral assessments is to clearly describe how each task, as outlined in the content specifications, will be assessed. To determine this, it is helpful to consider the cognitive complexity or demand of the content. This refers to the expected mental complexity a candidate requires to respond to a given item (Haladyna & Rodriguez, 2013). The most common and evidence-based approach to classifying cognitive complexity is the Bloom's Cognitive Taxonomy (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956), which in its revised version, consists of six levels organized according to increasing levels of cognitive complexity: remembering, understanding, application, analysis, evaluation, and creation (Anderson & Krathwohl, 2001). This taxonomy helps to determine the most appropriate method for assessing the content. The SMEs agreed that to ensure an oral assessment is the best method for assessing the content, it is necessary to go beyond the cognitive levels of "remembering" and "understanding" and to operate at the level of "application", "analysis", and "evaluation". These levels are also most aligned with the purpose of the oral assessments, which is to apply technical knowledge and make appropriate decisions.

Oral assessments, like most performance assessments, allow the measurement of process (how a task is completed) and product (the outcome of the task). The desire to assess both process and product is aligned with the use of a performance assessment for numerical standards and risk-based standards. In particular, since the purpose of the oral assessments is to measure application of technical knowledge and decision making, the focus of the items or questions

would be on assessing not only on whether the candidate correctly answered the question (product) but also on how they answered the question (process).

To encourage the SMEs to ensure the assessment context emulates the practice context as much as possible, the psychometric consultant presented the SMEs with two approaches in assessing the content: a scenario-based approach and a simulated-based approach. In the former approach the candidates would be presented with a series of scenarios with specific questions tied to the scenarios. In the latter approach, the candidate would be asked to respond to a simulated application that they would be expected to review. After a thorough discussion, the SMEs decided to adopt the scenario-based approach as they felt it was the most feasible way to assess the content since it would not be possible or fair to have candidates review a simulated application in a short amount of time. To ensure the assessment context is aligned with the CSAP practice context as much as possible, the SMEs agreed that scenarios will be developed by experienced and practicing APs who will be instructed to develop scenarios based on actual instances encountered by APs in their field of work.

Each question will be directly mapped on to the primary task category as indicated in the blueprint. For instance, in the numerical standards based oral assessment, there are 7 primary task categories. Therefore, in this oral assessment all 7 primary task categories will be assessed via a number of questions. The actual number of questions will be determined based on the weighting of the categories. Categories that are weighted the most will require more questions than categories that are weighted less. The number of scenarios presented to the candidates will vary from four to five, depending on how well the blueprint specifications are met.

To ensure valid, fair, and equitable measurement of the content for all individuals it is necessary to think about how the assessment process can be modified for candidates that request accommodations. In thinking about what kind of accommodations can be provided it is important to examine whether the accommodation will interfere with an accurate and valid measurement of the intended construct. Therefore it is necessary that each accommodation request is reviewed carefully to ensure that any anticipated interference with the measurement is minimized and that evidence is collected to ensure comparability between accommodated and unaccommodated assessments is valid. It is understood that the CSAP Society currently has a system in place in which requested accommodations are reviewed on a case-by-case basis.

It is currently estimated that each candidate may likely require 2 hours to complete the assessment. If candidates require more time, then they will need to provide documentation in support of this request during the registration process.

Scoring Specifications

Regarding scoring specifications the SMEs were instructed to think about how candidate responses will be scored, whether item scores will be combined into an overall score and what the score report will look like. To help with this process the SMEs were presented with two types of scoring approaches: analytic and holistic. In the analytic scoring method, a candidate's response is scored on several different components or elements that are aligned with expectations, whereas, in the holistic scoring method, a single score is generated based on an overall or holistic evaluation of whether the candidate's response is aligned with expectations or not. The SMEs decided that the analytic method would be most appropriate to use in scoring the oral assessment as it would allow them to see how candidates perform in each blueprint category. This scoring method would require the candidates to meet an established minimum level of proficiency for each category. The SMEs agreed that to pass the oral assessment the candidates would need to achieve an established minimum level in each blueprint category. The minimum level required to pass will be defined during the standard setting phase.

Next, the SMEs were asked to think about the scoring process. Although it is recommended that the scoring guide is developed during the item writing phase, it is important that at this stage there is a clear conception of the kinds of questions that will be asked, and what the candidates will be asked to do or how they will be expected to respond. Based on this information, ideal responses will need to clearly be delineated (i.e. what major points need to be included in the ideal response), as well as what aspects of the responses will be evaluated and how points will be assigned for each element. For instance, one of the SMEs mentioned that it would be important to assess the candidate's ability to know where to go to find information to respond to a given question. The SMEs provided additional aspects of responses that would need to be evaluated such, "Did the candidate end up with the right answer?", "Did the candidate stray from the correct answer?", "Did the candidate stray from the correct references or sources?", "Did the candidate include the correct references or sources?", and "Did the candidate follow the correct sequence of activities?". Based on the discussion with SMEs on the scoring process, the psychometric consultant recommended a checklist methodology rather than a scoring rubric to evaluate candidate responses, since the SMEs were more interested in examining whether a candidate met an expectation or not, rather than examining the degree to which an expectation was met or not. The latter is aligned with the use of a scoring rubric, which allow the evaluation of the extent to which criteria have been met (APA et al., 2014).

The SMEs indicated that there are four SMEs including themselves that will be able to take on the role of examiner. Training will be provided to all examiner by the psychometric consultant to ensure they are prepared to score candidate responses accurately. A training manual will also be provided. As part of the training the examiners will score sample responses and engage in discussion.

Finally, the SMEs were asked to envision the score report. It is recommended to do this at this stage as it would ensure alignment with the purpose of the assessment. The SMEs stated that only an overall pass/fail outcome should be reported on the score reports so that "no bias is introduced at the second attempt". However, the psychometric consultant recommended that it would be useful to report pass/fail information for each blueprint category, (especially for candidates that failed the assessment) so that it is clear on what areas candidates need to improve.

Assessment Administration Specifications

Testing administration specifications provide clear and detailed information on how the oral assessment will be administered. This includes information on "mode of test delivery, time limits, accommodation procedures, instructions, and materials provided to examiners and examinees, and procedures for monitoring test taking and ensuring test security" (AERA et al., 2014; p. 80).

Although the SMEs agree that the primary mode of test delivery will be oral-based, candidates will be provided with the list of questions that will be asked and some paper and pencil to write their responses for a given amount of time before beginning the oral assessment. This will likely help candidates manage their stress as it will provide an opportunity to collect their thoughts and prepare for oral presentation. This may also help non-native English speakers prepare for oral presentation of their responses.

As indicated in the format specifications, the SMEs estimated a two-hour duration for each candidate. For instance, the SMEs estimate that in the first hour candidates will be given the questions to prepare their responses. In the next 30 minutes, the oral assessment will be conducted. The last 30 minutes will be given to the evaluators to document findings, take a break, and then transition to the next candidate. In evaluating this timeline the psychometric consultant recommended 60 minutes for candidate preparation, 45 minutes for the oral assessment, and 15 minutes for evaluators to summarize results and prepare for the next candidate. This is a typical oral assessment administration process (Swanson, Norman, & Linn, 1995).

As mentioned in the format specifications, requested accommodations will be reviewed on a case-by-case basis. Accommodations that do not interfere with accurate and valid measurement of the construct of the oral assessment will likely be granted. The following accommodations will likely be provided to candidates after their documented need for the requested accommodation has been approved by the CSAP Society: an additional 30 minute time extension for the oral assessment, large font size and/or an assistant reader or scribe for visually challenged candidates, language interpreter for non-native English speakers, sign-language interpreter for hard-of-hearing or deaf candidates, and availability of extra breaks.

The oral assessment will be conducted on one candidate at a time. Upon entering the assessment site, the candidates will be asked to surrender their cell phones and other communication devices such as smart watches. Each candidate will be provided a reference list that can be used to determine which printed documents to bring to the evaluation, task lists for the evaluation and equations/tables. The oral assessment will be an open-book assessment in which candidates will have at their disposal all the necessary documents needed to respond to the questions (i.e., equation sheets, reference tables, reference guide, etc.). Candidates will be provided with a laptop, calculator, paper, and pencil. The oral assessment process will be audio recorded for quality control and training purposes, as well as in the event a re-scoring is required due to discrepancies between the scores assigned by the examiners.

Due to financial and logistic constraints, only one form of the risk-based standards assessment and one form of the numerical standards assessment will be developed and administered every year. Therefore, to ensure test security, the psychometric consultant recommended that each oral assessment be held on the same day at two consecutive time intervals. In this approach, oral assessments for each candidate will need to be conducted simultaneously. For instance, if there are only four examiners then only two candidates will be assessed at the same time by two examiners each, in separate rooms at the assessment site. If there are more than three candidates, the rest will need to be assessed at the second administration time interval which would be followed immediately after the end of the first administration. This approach will minimize leaking of assessment information from one candidate to the next. In the event there are more than four candidates then the others will be asked to wait in a waiting room to minimize interactions between candidates.

Finally, feedback from the candidates will be collected for improvement purposes. This information will be a critical component of the validation process.

Development of Questions and Scoring Methods

After finalization of the assessment blueprint, items or questions for the oral assessments can be developed. The question writing process is an intensive process that requires item developers to adhere to strict guidelines on item writing and to ensure each item maps directly onto a specific primary task category. The psychometric consultant provided an item development tutorial to 4 numerical SMEs and 4 risk-based SMEs, separately. In the tutorial, the psychometric consultant started by providing information to the SMEs on the purpose of the oral assessments, the blueprint for the oral assessment, and what the oral assessment will not be assessing (i.e. technical knowledge and content already assessed in the regulatory exam). Next he talked about how the oral assessment will be administered.

Due to time limitations (i.e. the oral assessment must be ready for administration by September 2023), there may not be sufficient time to develop enough new questions and therefore it became necessary to select and adapt scenario-based multiple choice questions from the

existing risk-based and numerical item banks for oral assessment use. Therefore the psychometric consultant facilitated a discussion on how best to select and adapt existing questions from the item banks to use for the oral assessment, through the examination of four key questions: 1) What criteria will be used to select scenarios? 2) What features should an appropriate scenario have? 3) What kind of questions to ask? 4) What are the expected responses? Some guidelines on how to develop realistic scenarios was also provided. The SMEs decided that the chairperson (the same two SMEs who participated in the practice analysis and blueprint development process) will be selecting existing items from the bank for adaptation. The chairperson will then assign items for adaptation to the team. To provide a framework or backbone for the development of the questions under each primary task category, the psychometric consultant introduced the concept of item shells (Haladyna and Shindoll 1989). Item shells force the developer to think about the features of questions belonging to each primary task category. To help develop item shells, the SMEs were encouraged to consider these key questions: What features do questions belonging to the same primary task category share? What features are distinct between the primary task categories? This item writing approach has been successfully used in credentialing assessments in the health care context and are recommended for developing performance-based tasks and questions (Haladyna and Shindoll 1989).

As mentioned earlier, oral assessments are judge-mediated assessments (Stone et al., 2008) in which candidate responses are interpreted and scored by a panel of judges or examiners. Therefore the SMEs were asked to think about expected responses to questions and how the questions will be scored. To help the SMEs think about the scoring process the psychometric consultant conducted a separate tutorial to provide information, resources, and guidance on appropriate selection of scoring methods.

To ensure oral assessment scores are valid, reliable, and fair, it is necessary to adhere to quality assurance processes that include “documentation of overall guidelines to be utilized for a particular assessment, clearly defined scoring rubrics, and a well-articulated scoring design (including length and content of scorer training, the frequency and types of scorer monitoring, roles and responsibilities of various stakeholders, etc.). Scorer training focuses on improving the consistency of scoring and reducing rater bias. Quality control procedures include specific techniques such as rater calibration (assessing and certifying raters before scoring begins), back-reading (having expert scorers rescore samples of scored papers), validity scoring (inserting previously scored papers that represent particular score points into the scoring process) and monitoring statistics” (e.g., agreement rates and related statistics for papers scored by two or more raters; Way & Gialluca, 2017; p. 107).

The purpose of quality control processes for question writing and scoring is to minimize threats to validity. Two common threats to validity are identified: construct irrelevant variance (CIV) and construct underrepresentation. “CIV occurs when the examination assesses one or more unintended constructs in addition to the construct it was designed to measure (Messick, 1989).

Some common CIVs in the oral assessment context are providing long and elaborate responses, and speaking comfortably without any hesitation. Extensive review of the questions and scoring process is necessary to ensure that these CIVs are minimized as much as possible. Construct underrepresentation occurs when the either the assessment does not cover the necessary domain of the content or the scoring process does not accurately capture the responses not identified in the rubric as being correct. Construct underrepresentation in the question writing process is typically minimized by ensuring that the assessment meets blueprint specifications. This is done by ensuring the proportion of questions assessing each competency identified in the blueprint matches the blueprint specifications. Construct underrepresentation in the scoring process is typically minimized by having two independent examiners scoring each candidate's responses and flagging scores that are discrepant (i.e examiners provide different scores for a given candidate). All of these approaches were implemented in the CSAP Oral Assessments.

There were 4 SMEs in each of the oral assessment group that participated in writing questions and expected responses for the oral assessments. The chair of each group assigned each of the SMEs two questions to write, based on the content specifications in the blueprint. The goal was to create 20 questions for the numerical oral assessment and 20 questions for the risk-based oral assessment. The SMEs were encouraged to revise and adapt existing multiple-choice questions (which were used in the previous exams) into appropriate scenarios and questions for the oral assessment. This was done so that enough questions could be developed in time for the September 2023 administration of the oral assessments, since it would take longer for the SMEs to develop new questions than to revise and adapt existing ones. The psychometric consultant advised the SMEs to develop 4 to 5 scenarios that consisted of no more than 4 to 5 questions for each scenario. The goal was to produce 20 questions, as opposed to 20 scenarios. The psychometric consultant also advised the SMEs to describe the expected response clearly and concisely for each question as well as the number of points needed to obtain full marks as well as whether and how many partial marks would be provided. Each question was created by the SMEs independently. Each question was then reviewed by one of the SMEs (questions to review were assigned by the chair). In addition to reviewing the content of the questions, the SMEs were instructed to use the question review checklist for quality control purposes. Once the questions and expected responses were peer-reviewed the psychometric consultant reviewed all the questions and expected responses from a psychometric perspective. Feedback from the psychometric consultant was provided at the standard setting session.

Assessment Assembly

The 2023 Numerical Standards Oral Assessment

The 2023 numerical standards oral assessment consisted of 13 scenarios with 1-7 questions each, contributing to a total of 52 scenario-based questions. Each question was worth between 0.5 to 1.5 points each. The cut score was calculated by averaging the expected scores from each rater or panelist for each question and then summing up the average expected means across

the questions (Hambleton & Plake, 1995). The questions were developed and reviewed by a group of six subject matter experts (SMEs) using a structured item development and review process.

The items map to the following content areas outlined in the Exam Syllabus contained in the "Examination Guide for Exam Candidates" as shown in Table 7.

Table 7: *Exam coverage of syllabus content areas*

Content Area	**Number of points on 2023 Numerical Exam**	**Percentage of points on 2023 Numerical Exam**	**(Approximate) Percentage of Exam Items Indicated in Exam Syllabus**
Historical and Visual Site Information	4	10.8%	11%
Assessment of Affected Media and Migration Pathway	6	16.2%	16%
Contaminant Characteristics	5	13.5%	15%
Investigation Approach and Methods	6	16.2%	16%
Data Synthesis and Interpretation	8	21.6%	21%
Risk Assessment Principles and Screening	3	6.8%	7%
Remediation Design, Implementation and Confirmation	6	14.9%	14%
TOTAL	**38**	**100%**	**100%**

The 2023 Risk-based standards Oral Assessment

The Risk-based exam consists of scenario-based questions that were worth 2 points each. The cut score was calculated by averaging the expected scores from each rater or panelist for each question and then summing up the average expected means across the questions (Hambleton & Plake, 1995). The questions were developed and reviewed by a group of six subject matter

experts (SMEs) using a structured item development and review process. The 2023 oral assessment consisted of 20 scenarios, which contained 1-4 questions in each scenario, contributing to a total of 42 scenario-based questions. Each question was worth between 0.25 to 1 point each.

The items map to the following content areas outlined in the Exam Syllabus contained in the "Examination Guide for Exam Candidates" as shown in Table 8:

Table 8: *Exam coverage of syllabus content areas*

Content Area	**Number of Scenarios on 2023 Risk Exam**	**Percentage of Scenarios on 2023 Risk Exam**	**(Approximate) Percentage of Exam Items Indicated in Exam Syllabus**
Ecological Risk Assessment	10	50%	50%
Problem Formulation	3	15%	15%
Exposure Assessment	3	15%	12.5%
Effects Assessment	2	10%	12.5%
Risk Characterization, Uncertainty Analysis, and Submission Eligibility	2	10%	10%
Human Health Risk Assessment	10	50%	50%
Problem Formulation	3	15%	15%
Exposure Assessment	3	15%	15%
Toxicity Assessment	2	10%	10%
Risk Characterization, Uncertainty Analysis, and Submission Eligibility	2	10%	10%
TOTAL	**20**	**100%**	**100%**

Standard Setting

Standard setting is defined as "the proper following of a prescribed, rational system of rules or procedures resulting in the assignment of a number to differentiate between two or more states or degrees of performance' (e.g. pass/fail)" (Cizek, 2011, p. 50). In practical terms it is a scientific process of determining an appropriate and defensible minimal score required to pass an assessment. In standard setting it is important to distinguish between a standard and a cut score. "A standard is a qualitative description of a level of performance and can be viewed as a conceptual definition of competence, whereas a cut-score or passing score corresponds to a

number that reflects this standard and can be viewed as an operational definition of competence" (De Champlain, 2019; p. 348). Since standard setting is a judgemental process it can become inherently subjective. Therefore it is important to gather validity evidence at every stage of standard setting to ensure that the process through which a cut score has been established is valid, reliable, and fair. This requires careful consideration of 1) recruitment and selection of appropriate judges who will be involved in setting the standard and cut score, 2) selection and then implementation of an appropriate standard setting method, 3) training of judges to ensure they know how to use the method to set an appropriate standard and cut score, and then 4) monitoring their work and conducting analyses to evaluate the reliability, validity, and fairness of the cut score. "What is ultimately of importance with any standard setting exercise is the extent to which a process is systematically adhered to and can be defended using a number of evidential sources (De Champlain, 2019; p. 356). Specifically, evidence for procedural validity (i.e. how the process was conducted) internal validity (i.e. how precise the cut score is), and external validity (i.e. how the cut score impacts the pass rate) is collected to ensure the standard setting process is defensible.

Virtual standard setting sessions were conducted for each of the two oral assessment group separately. The psychometric consultant began the session by providing some feedback on the questions as well as the scoring process. The psychometric consultant provided the following feedback for both oral assessment groups:

- Overall, excellent work!
- Please make sure there are no questions assessing knowledge
- Some of the scenarios only had 1 question
- Do all questions in a given scenario assess the same blueprint category? If not, please review and specify the appropriate category for each question.
- Each question should measure a specific blueprint category – not scenario
- Instead of having a unique item ID for scenario, please provide a unique item ID for each question.
- Please make sure that the blueprint percentages are met
    - We need to make sure we have sufficiently large number of questions in each category in order to have high level of measurement reliability
- Scoring guide needs to be clearly defined and developed before we can do standard setting
- Think about what types of questions are asked in each blueprint category
    - E.g. What kinds of questions are asked under "Contaminant Characteristics"?
- This will allow us to develop a generic scoring guide that can guide the development of a specific scoring guide for each question
- Scoring rules must be clearly defined and consistent across questions
    - E.g. What are the total possible scores for a question asking about references?
    - E.g. What are total possible scores for a calculation question?
    - Any part marks?

The psychometric consultant also shared a draft of the scoring guide. This stimulated a discussion among the SMEs on how best to score the responses.

The psychometric consultant then proceeded to provide a training on the standard setting process. Since oral assessments can be considered a type of performance-based assessment it was necessary to first focus on compensatory vs. conjunctive approaches to standard setting. In compensatory standard setting, a cut score is based on overall performance on the assessment, rather than how a candidate performed at each blueprint category. In conjunctive standards setting, a cut score is set at the blueprint category level. Therefore, in order to pass the assessment, a minimum performance level must be obtained at each blueprint category. The pros and cons for each of these standard setting approaches was discussed. The psychometric consultant indicated that in order for a cut score to be reliable, valid, and fair, it needed to be based on a sufficiently large number of questions. Since there are likely to be very few questions at each blueprint category it would not be advisable to set separate cut scores for each category. After much discussion the numerical standards oral assessment group decided to adopt a compensatory approach such that one cut score would be set at the total assessment score level. The risk-based oral assessment group decided to set two cut scores - one for the ecological risk assessment level and one for the human health risk assessment level.

Next, the psychometric consultant talked about test-centred standard setting methods vs. examinee-centered standard setting methods. In test-centered standard setting methods, SMEs review each question and determine how difficult it would be for a minimally qualified candidate. In examinee-centered standard setting methods, SMEs review actual performance of minimally qualified candidates to derive a cut score. Since the latter methods require a sufficiently large sample of minimally qualified candidates, examinee-centered methods would not be appropriate for the CSAP Society since their exams are all small-scale or consist of very low candidate volumes. Therefore the psychometric consultant advised the SMEs to use a test-centered standard setting method for determining the cut score for the oral assessments. Currently, the modified Angoff method, which is a test-centered standard setting method, is used to set the cut score for the knowledge-based regulatory exam. For oral assessments, like most performance-based assessments, the Extended Angoff method is recommended (Hambleton & Plake, 1995).

In the Extended Angoff method, SMEs are instructed to first clearly define the minimally qualified candidate and then to go through each question independently and determine what score the minimally qualified candidate will likely achieve. In other words, they were instructed to provide an expected score (i.e. extended Angoff rating) for each question. MCCs are hypothetical candidates who are on the borderline between competence and non-competence, but who would just barely meet the performance level required for competence.

Once the MCC has been defined and the SMEs express a common understanding of the characteristics of the MCC derived through group discussion, the SMEs are asked to consider a group of 100 MCCs and estimate the score 100 MCCs would obtain on each question on the

assessment. In other words, they are asked to independently come up with the expected score the MCC will achieve on each question. This constitutes Round 1. Upon completion the ratings are compiled by the facilitator and psychometrician and items that have discrepancies greater than 0.5 points are flagged for group discussion. During group discussion SMEs discuss their rationales for their ratings. The purpose of group discussion is to encourage consensus by ensuring all SMEs perceive the MCC in the same way. After discussion, panelists are invited to consider revising their ratings but are not required to make any changes. This constitutes Round 2.

Standard Setting the 2023 Numerical-based oral assessment

Definition of Minimally Competent Candidate

The definition of the MCC for this assessment was defined during previous years' exam development in the following way:

The minimally competent candidate is someone who:

- At least 10 years of relevant documented experience (RDE)
    - Work experience in fields of practice applying biological, toxicological, geoscience, engineering, or agrology principles.
- At least 8 years of direct documented experience (DDE)
    - Experience in the field of contaminated sites investigation, risk assessment, management and/or remediation.
- At least 4 years of decision-making level DDE
    - Providing direction and senior review of scientific and technical tasks necessary to complete investigations, risk assessment, management and remediation at contaminated sites.
- Understands how to apply the numerical standards of the Contaminated Sites Regulation (CSR)
- Understands how to apply combined aspects of ecology, toxicology and environmental chemistry for the review of human health and ecological risk assessments.
- Has a general understanding of related areas such as basic contaminant transport in various media.

This definition of the MCC was reviewed during the initial training session, and informed by a practice analysis survey that was administered to numerical standards Approved Professionals in the field. The practice analysis survey for numerical standards was completed by six participants (response rate of 100%) during the Spring of 2023. Sixty-seven percent of participants identified as numerical APs, while the rest identified as both numerical and numerical standards APs.

The following table summarizes the findings from the practice analysis survey. They were asked to rate the degree to which each knowledge statement is needed to successfully accomplish the

task of a numerical standards Approved Professional. Each knowledge statement, which identifies what the MCC knows and can do was rated on a scale from 1 (Not Important) to 5 (Critically Important). The knowledge statements in Table 1 are organized by most important to least important. This table provided SMEs with a good understanding of what the MCC is expected to know and do.

Table 9: Expected knowledge of Minimally Competent Candidates

Knowledge Statement	Average Rating
1. Know what to consider an APEC, and how to use historical data to get there.	5
2. Know what PCOCs to assess for.	5
5. Understand the process for submission of legal instruments.	5
16. Regulatory requirements and standards.	5
18. Understand the investigation, remediation and characteristics of common contaminants.	5
32. Know potential contaminants of concern associated with most common land usage.	5
33. Know how to determine what processes/land uses are APECs and their related PCOCs.	5
34. Determine groundwater flow direction using representative wells.	5
37. Know when to go to an expert.	5
43. Verify proper sampling methods, field screening, laboratory tests, and holding times.	5
46. Develop a conceptual site model.	5
48. Identify data gaps and their significance.	5
49. Determine what standards apply to a site.	5
53. Identify basic chemical properties.	5
65. Be aware of own limitations and limitations of reliance on data for their site.	5
66. Be able to back out of situations and know who to ask if necessary.	5

72. Know when judgment is allowed.	5
74. Review own work before submitting it.	5
3. How to properly assess a site (drilling, sampling, lab analysis)	4.75
4. Be able to source information from the CSR and Guidance documents.	4.75
6. Understand "101" level hydrogeology, chemistry, and biological "specialist" concepts.	4.75
9. Historical, current and potential future site land use Information.	4.75
11. Contaminant identification/characterization.	4.75
17. Understand the regulations and how they are implemented technically.	4.75
39. Understand what a hydraulic conductivity value means.	4.75
42. Find physical and chemical properties of contaminants.	4.75
50. Use Stage 1 PSI conclusions to plan a Stage 2 investigation.	4.75
51. Use Stage 2 PSI results to plan a DSI program.	4.75
52. Develop a sampling plan.	4.75
64. Know standard investigative procedures and remedial approaches.	4.75
67. Recognize gray areas in terms of data interpretation.	4.75
70. Have been exposed to most situations but may need to look up things.	4.75

After the MCC discussion, SMEs felt they had a good understanding of who the MCC is, and what he/she knows and can do. Their feedback was collected after the standard setting process.

The Rating Process

The standard setting exercise began with an orientation delivered by the psychometrician, in which he talked about how to implement the Extended Angoff method, including how to generate expected scores for each question (i.e. Angoff ratings) and the factors to keep in mind when making ratings. Prior to beginning the rating process, two sample questions were presented, rated and discussed by the group as a means of "calibrating" and again discussing the definition of the MCC prior to rating the selected items. Upon completion of this calibration exercise and before beginning the rating process the SMEs were asked if they had any questions and were reminded to keep the written definition of the MCC beside them and at the top of their mind while completing the exam and rating the items. The SMEs indicated that they understood the task and were ready to begin rating.

All SMEs then went though each of the questions and independently provided an expected score (i.e. extended Angoff rating) for each question. Items that had score discrepancies of greater than 0.5 were flagged for discussion. Three questions had discrepancies greater than 0.5 (minimum discrepancy = 1.0, maximum discrepancy = 1.5). The SMEs scheduled a separate meeting time to discuss the discrepancies together. Discrepancies in Round 2 ratings ranged from 0 to 0.5.

Results

The Numerical standards oral assessment consists of 52 scenario-based questions from 13 scenarios that were worth between 0.5 to 1.5 points each. The cut score was calculated by averaging the expected scores from each rater or panelist for each question and then summing up the average expected means across the questions, as recommended by Hambleton & Plake, 1995.

The total expected score (i.e. expected number of points) out of 38 points provided by each SME, along with the group mean, is presented in the table below. The total expected score was calculated by summing the expected scores across the 52 questions.

Table 10: Average SME ratings

	SME 1	SME 2	SME 3	SME 4	Group Mean
Round 1	22	18	22	24	21
Round 2	23	21	24	27	24

The summary statistics are provided the table below. The reliability of ratings across the SMEs is 0.96. Although this reliability value is satisfactory (Cicchetti, 1994) the result should be cautiously interpreted because the items were rated by a relatively small group of raters.

Table 11: Summary statistics

Number of scenarios	13
Number of scenario-based questions	52
Maximum number of points	38
Number of SMEs	4
Expected number of points out of 38	24
Expected percent correct	63%
Median rating	1.5
Range of ratings	0.5 – 4.5
Standard Error of judgements (SEJ)	0.03
Standard Error of item means (SEI)	0.13
Reliability	0.96
Standard Error of Measurement	0.12

The mean Angoff rating and corresponding raw score is presented below in Table 3, including ratings and corresponding raw scores that are 1, 2, and 3 standard errors of judgement above and below the mean Angoff rating. The standard error of judgements (SEJ) allows one to determine the validity of the cut score recommendations that are based on just one panel of judges (Tannenbaum & Cho, 2014). SEJ "indicates how close the mean cut score would likely be to the current mean cut score for other panels of SMEs similar in composition and experience to the current panel and similarly trained in the same standard-setting methods. A comparable panel's cut score would be within 1 SEJ of the current mean cut score 68% of the time, within 2

SEJs 96% of the time and within 3 SEJs 99% of the time" (Tannenbaum & Katz, 2008; p. 12). An acceptable SEJ is one that is less than half the standard error of measurement (SEM) (Cohen, Kane, & Crooks, 1999). As indicated in Table 12 above this is indeed the case. Since the SEJ is quite small it does not change the cut score even after +/- 3 SEJ.

Table 12: Angoff ratings +/- 3 Standard Error of Judgement

	-3 SEJ	-2 SEJ	-1 SEJ	Angoff	+1 SEJ	+2 SEJ	+3 SEJ
Percentage	63%	63%	63%	63%	63%	63%	63%
Raw score out of 38 points	24	24	24	24	24	24	24

The standard error of measurement (SEM), which is reported the above table, provides an estimation of the amount of error there is in using the established cut score to make pass-fail decisions. Specifically, the SEM can be used to compute confidence intervals around the cut score to provide an idea of how much error there is in using the cut score. If necessary, this information can be used to adjust the cut score.

A 99% confidence interval lies within the area bounded by 3 SEMs below and 3 SEMs above the recommended cut score, a 95% confidence interval is bounded by 2 SEMs below and above the recommended cut score, and a 68% confidence interval is bounded by 1 SEM below and above the recommended cut score. These values are shown in the table below, expressed both as a percentage of the total score and as a raw count of items out of 38. Due to the very small SEM value, the cut score does not change much even after adding and subtracting 3 SEM values.

Table 13: Angoff ratings +/- 3 Standard Error of Measurement

	-3 SEM	-2 SEM	-1 SEM	Angoff	+1 SEM	+2 SEM	+3 SEM
Percentage	63%	63%	63%	63%	63%	63%	63%
Raw score out of 38 pts	24	24	24	24	24	24	24

After being presented the expected score of each SME, including this score minus the standard error of measurement (24 – 0.12 = 23.88) and this score plus the standard error of measurement (24 + 0.12 = 24.12), the SMEs discussed these statistics. They also considered the cut score of previous exams. One factor the SMEs considered was the fact that candidates will not have access to internet and online resources but will only be allowed to bring hard copies printed from online resources when they write the exam. This may likely slow them down as they flip through pages of paper to locate pieces of information to answer each item. However, after much deliberation the SMEs decided to keep the cut score at 24.

Here are all the comments regarding the cut score that SMEs provided on their evaluation forms:

- “I thought that the cut score was reasonable.”
- “If I recall correctly the cut score is a bit lower than it has been for many of the previous exams using the old multiple choice method suggesting it is a harder exam. The questions don't seem harder than previous years but the new oral assessment process may add some difficulty”.
- “I think it is a fair cut-score though I was expecting a higher cut score than 63% but if we are all within that range I am comfortable with it”

After the discussion the SMEs were unanimous in recommending a final cut score of 24 correct items out of 38 points. However, this is not the final cut score that was used for this assessment. The final cut score was recalculated and reviewed by the committee after removal of some poorly functioning questions and/or questions for which there were valid comments made by candidates.

Table 14: Standard setting session feedback

		Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	The standard setting training was helpful in providing me with a clear understanding of how to perform the task.				3	
2	The questions provide valid and fair opportunities to demonstrate CSAP knowledge, skills, and competencies.				1	2
3	I was able to conceptualize the minimally qualified candidate.				1	2
4	I rated each question based on how the minimally qualified candidate would perform.				1	2
5	I had sufficient time to complete my ratings.				2	1
6	The group discussions were helpful.				1	2
7	I am confident in the ratings I provided.				2	1
8	I am confident in the final ratings provided by the group.				1	2
9	I am confident in the method used to determine the cut score for this assessment.				2	1
10	The established cut score meets my expectations.			1	2	

Standard Setting the 2023 Risk-based standards oral assessment

Definition of Minimally Competent Candidate

The definition of the MCC for this assessment was defined during previous years' exam development in the following way:

The minimally competent candidate is someone who:

- At least 10 years of relevant documented experience (RDE)
    - Work experience in fields of practice applying biological, toxicological, geoscience, engineering, or agrology principles.
- At least 8 years of direct documented experience (DDE)
    - Experience in the field of contaminated sites investigation, risk assessment, management and/or remediation.
- Understands how to apply the risk-based standards of the Contaminated Sites Regulation (CSR)
- Understands how to apply combined aspects of ecology, toxicology and environmental chemistry for the review of human health and ecological risk assessments
- Has a general understanding of related areas such as basic contaminant transport in various media.

This definition of the MCC was reviewed during the initial training session, and informed by a practice analysis survey that was administered to risk-based standards Approved Professionals in the field. The practice analysis survey for risk-based standards was completed by six participants (response rate of 100%) during the Spring of 2023. Sixty-seven percent of participants identified as risk-based CSAPs, while the rest identified as both risk-based and numerical standards CSAPs.

The following table summarizes the findings from the practice analysis survey. They were asked to rate the degree to which each knowledge statement is needed to successfully accomplish the task of a risk-based standards Approved Professional. Each knowledge statement, which identifies what the MCC knows and can do was rated on a scale from 1 (Not Important) to 5 (Critically Important). The knowledge statements in Table 1 are organized by most important to least important. This table provided SMEs with a good understanding of what the MCC is expected to know and do.

Table 15: Expected knowledge of Minimally Competent Candidates

Knowledge Statement	Average Rating
1. Know what “arm’s length review” is.	5
15. Describe the difference between an HQ and ILCR.	5
2. Be able to review risk assessment and identify where rationale is inadequate.	5
21. Determine if certain risk assessments are worthy of being submitted under Protocol 6.	5
26. Determine if risk assessment is worthy of submission under protocol	5
3. Know key ministry guideline documents.	5
11. Use fundamental dose equations (e.g., Health Canada, ORNL).	4.8
16. Characterize risk; calculate risk estimates.	4.8
23. Use fundamental dose equations (e.g., Health Canada, ORNL).	4.75
13. Use and convert these values in order to calculate human health risks (HQ and ILCR).	4.6
14. Know the basic standard measures of toxicity to ecological receptors (e.g. LC50, EC20, IC20, NOEC, LOEC, NOAEL, LOAEL).	4.6
17. Understand how exposure concentrations should be selected for different receptors.	4.6
18. Pick the right variable for an equation.	4.6
25. Able to apply regulatory requirements.	4.6
22. Explain fundamental ecological/biological concepts relating to quantitative methods used in ecological risk assessment.	4.5
24. Explain the different types of human health toxicological reference values, i.e. RFD, tolerable concentration, slope factor, unit risk (i.e. what they represent).	4.5
12. Explain the different types of human health toxicological reference values, i.e. RFD, tolerable concentration, slope factor, unit risk (i.e. what they represent).	4.4

10. Explain fundamental ecological/biological concepts relating to quantitative methods used in ecological risk assessment.	4.2
6. Make submissions independently under Protocol 6 (administrative work, flawless submissions).	4.2
19. Pick a TRV from a source.	4
4. Explain all decisions; provide rationale behind things.	4
20. Know how that TRV was calculated, and how to leverage that information in rationale.	3.8
7. Define a chemical carcinogenic substance and how it differs in different areas.	3.8
9. Explain fundamental toxicological principles (adsorption, transformation, target tissue effects, excretion).	3.8
5. Explain development of federal standards and provincial guidelines; know the history.	3.6
8. Describe contaminant chemistry and its influence on fate and transport.	3.2

This information was combined with the following list from a previous discussion on MCC characteristics to inform the definition of the MCC:

What will the MQC be able to do quite well?

- Different types of human health, toxicological reference values, tolerable intakes and be able to convert values to calculate risks, combining exposures, dosages, etc. to estimate human health risk
- Fundamental understanding of science behind the work we do
- Basic understanding of regulatory setting – what we can and cannot get away with
- Know where to look for info that's pertinent to requirements
- Screening, risk calculations, TRV, exposure calculations, problem formulations,
- General understanding of risk assessment theory
- Understands positions of other agencies
- Understanding and able to apply regulatory requirements

- Know what arms-length review is
- Determine if risk assessment is worthy of submission under protocol 6

What will the MQC find challenging?

- Deriving these values is not expected to do
- Not have full understanding of all guidance documents and their evolution
- Coming up with answers on conflicting guidance
- De novo toxicity reference values derivations
- Applications of sub chronic human health TRVs
- Bio accessibility
- Details on rationale and specific input parameters and TRVs and specific guidance of other agencies
- Complex risk assessments

After the MCC discussion, SMEs felt they had a good understanding of who the MCC is, and what he/she knows and can do. Their feedback was collected after the standard setting process.

The Rating Process

The standard setting exercise began with an orientation delivered by the psychometrician, in which he talked about how to implement the Extended Angoff method, including how to generate expected scores for each question (i.e. Angoff ratings) and the factors to keep in mind when making ratings. Prior to beginning the rating process, two sample questions were presented, rated and discussed by the group as a means of "calibrating" and again discussing the definition of the MCC prior to rating the selected items. Upon completion of this calibration exercise and before beginning the rating process the SMEs were asked if they had any questions and were reminded to keep the written definition of the MCC beside them and at the top of their mind while completing the exam and rating the items. The SMEs indicated that they understood the task and were ready to begin rating.

As mentioned previously, only four SMEs went though each of the 42 questions and collectively provided an expected score (i.e. extended Angoff rating) for each question. As a result of providing expected scores collectively, rather than independently (as instructed by the psychometrician), there was no variability of scores across raters, and therefore no questions were flagged for discussion in round 2. Therefore, the psychometrician decided to treat these four SMEs as one SME and to request to the CSAP Society to recruit another group of SMEs to independently rate the questions. Two more SMEs were then recruited to provide Angoff ratings for the set of questions independently. These two SMEs provided round 1 ratings which were compared to the previous group's ratings to determine rating discrepancies. Items that had score discrepancies of greater than 0.5 were flagged for discussion. Eight questions out of the 42 questions had discrepancies greater than 0.5 (minimum discrepancy = 0.75, maximum

discrepancy = 1.25). The SMEs scheduled a separate meeting time to discuss the discrepancies together. Discrepancies in Round 2 ratings ranged from 0 to 0.5.

Results

The Risk-based standards oral assessment consists of 42 scenario-based questions from 20 scenarios that were worth between 0.25 to 1 point each. The cut score was calculated by averaging the expected scores from each rater or panelist for each question and then summing up the average expected means across the questions, as recommended by Hambleton & Plake, 1995.

The total expected score (i.e. expected number of points) out of 40 points provided by each SME, along with the group mean, is presented in the table below. The total expected score was calculated by summing the expected scores across the 42 questions.

Table 16: Average SME ratings

	SME 1 group	SME 2	SME 3	Group Mean
Round 1	31	32	29	30
Round 2	31	29	28	29

The summary statistics are provided the table below. The reliability of ratings across the SMEs of 0.82. Although this reliability value is satisfactory (Cicchetti, 1994) the result should be cautiously interpreted because the items were rated by a relatively small group of raters.

Table 17: Summary statistics

Number of scenarios	20
Number of scenario-based questions	42
Maximum number of points	40
Number of SMEs	3
Expected number of points out of 40	29
Expected percent correct	73.21%
Median rating	0.75
Range of ratings	0.5 - 2
Standard Error of judgements (SEJ)	0.017
Standard Error of item means (SEI)	0.052
Reliability	0.82
Standard Error of Measurement	0.12

The mean Angoff rating and corresponding raw score is presented below in Table 18, including ratings and corresponding raw scores that are 1, 2, and 3 standard errors of judgement above and below the mean Angoff rating. The standard error of judgements (SEJ) allows one to determine the validity of the cut score recommendations that are based on just one panel of judges (Tannenbaum & Cho, 2014). SEJ "indicates how close the mean cut score would likely be to the current mean cut score for other panels of SMEs similar in composition and experience to the current panel and similarly trained in the same standard-setting methods. A comparable panel's cut score would be within 1 SEJ of the current mean cut score 68% of the time, within 2 SEJs 96% of the time and within 3 SEJs 99% of the time" (Tannenbaum & Katz, 2008; p. 12). An acceptable SEJ is one that is less than half the standard error of measurement (SEM) (Cohen, Kane, & Crooks, 1999). As indicated in the above table this is indeed the case.

Table 18: Angoff ratings +/- 3 Standard Error of Judgement

	-3 SEJ	-2 SEJ	-1 SEJ	Angoff	+1 SEJ	+2 SEJ	+3 SEJ
Percentage	73.01	73.13	73.15	73.21	73.25	73.28	73.33
Raw score out of 40 points	29.23	29.25	29.26	29.28	29.30	29.31	29.33

The standard error of measurement (SEM), which is reported in Table 4 above, provides an estimation of the amount of error there is in using the established cut score to make pass-fail decisions. Specifically, the SEM can be used to compute confidence intervals around the cut score to provide an idea of how much error there is in using the cut score. If necessary, this information can be used to adjust the cut score.

A 99% confidence interval lies within the area bounded by 3 SEMs below and 3 SEMs above the recommended cut score, a 95% confidence interval is bounded by 2 SEMs below and above the recommended cut score, and a 68% confidence interval is bounded by 1 SEM below and above the recommended cut score. These values are shown in the table below, expressed both as a percentage of the total score and as a raw count of points out of 40.

Table 19: Angoff ratings +/- 3 Standard Error of Measurement

	-3 SEM	-2 SEM	-1 SEM	Angoff	+1 SEM	+2 SEM	+3 SEM
Percentage	72.30	72.60	72.90	73.21	73.50	73.80	74.10
Raw score out of 40 pts	28.92	29.04	29.16	29.28	29.40	29.52	29.64

After being presented the expected score of each SME, including this score minus the standard error of measurement (29 – 0.12 = 29.16) and this score plus the standard error of measurement (29 + 0.12 = 29.40), the SMEs discussed these statistics. They also considered the cut score of previous exams. Another factor the SMEs considered was the fact that candidates will not have access to internet and online resources but will only be allowed to bring hard copies printed from online resources when they write the exam. This may likely slow them down as they flip through pages of paper to locate pieces of information to answer each item. However, after much deliberation the SMEs decided to keep the cut score at 29.

Here are all the comments regarding the cut score that SMEs provided on their evaluation forms:

- “It seems reasonable”.

- “Cut score is fair”
- “Cut score is fair and consistent with past years for the multiple choice exam”.

After the discussion the SMEs were unanimous in recommending a final cut score of 29 correct items out of 40 points. However, this is not the final cut score that was used for this assessment. The final cut score was recalculated and reviewed by the committee after removal of some poorly functioning questions and/or questions for which there were valid comments made by candidates.

Table 20: Standard setting session feedback

		Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	The standard setting training was helpful in providing me with a clear understanding of how to perform the task.				2	
2	The questions provide valid and fair opportunities to demonstrate CSAP knowledge, skills, and competencies.				4	
3	I was able to conceptualize the minimally qualified candidate.				3	1
4	I rated each question based on how the minimally qualified candidate would perform.				2	1
5	I had sufficient time to complete my ratings.				2	2
6	The group discussions were helpful.				3	
7	I am confident in the ratings I provided.				4	
8	I am confident in the final ratings provided by the group.				4	
9	I am confident in the method used to determine the cut score for this assessment.				2	
10	The established cut score meets my expectations.				2	

Assessment Administration

Administration of the 2023 Numerical standards Oral Assessment

On September 18, the psychometric consultant scheduled a virtual meeting with the seven examiners to go over the scoring process and to address any questions or concerns. It must be noted that the seven examiners were involved with the development and standard setting of the risk-based and numerical-based oral assessments. The cut score of the oral assessments was not revealed to the examiners in order to minimize any potential bias in scoring. After this meeting the examiners informed the psychometric consultant that they understood the scoring process and were ready to score responses on assessment day.

The numerical-based assessment was administered on September 19, 2023. Five candidates attended the assessment venue to write the assessment. Four examiners were in attendance. Two candidates were assessed simultaneously by a two separate panels of examiners.

The candidate to be assessed was directed into a room and were provided with instructions on the assessment process, starting with 90 minutes to prepare their responses to the assessment questions, followed by a 90 minute oral assessment in which they convey their responses to the panel of examiners as each question is asked. The instructions were provided by Nelly Pomarosa. After ensuring the candidate did not have any further questions, 90 minutes were provided to the candidate to begin reading the assessment questions and to prepare their responses using the paper and pencil provided to them. The candidate was also provided with an opportunity to note down any question-specific feedback, while responding to the questions. Nelly checked in with the candidate every 30 minutes and indicated how much time was remaining. Upon completion of the 90 minutes, the candidate was provided 15 minutes for a bio-break and then was directed to the oral assessment room. Two examiners were present in the assessment room. The 90 minute oral assessment component started with the examiners providing instructions on how to proceed. The candidate was also informed that the oral assessment component will be audio-recorded. After ensuring the candidate did not have any questions, the examiners took turns reading each scenario and accompanying question(s). The candidate provided their responses orally while both examiners used the examiner score sheet to score the candidate's response to each question. Upon completion of the oral assessment, the candidate was asked to complete an evaluation form before leaving the assessment site.

Administration of the 2023 Risk-based standards Oral Assessment

The risk-based assessment was administered on September 20, 2023. Three candidates attended the assessment venue to write the assessment. Three examiners were in attendance. Since two examiners were required to administer the assessment and score responses for each candidate, each candidate was provided with an assessment time that they were asked to adhere to. Upon arrival, each candidate was requested to surrender their cell phone and any other communication devices such as smartwatches. The candidates were then directed into separate waiting rooms. Each candidate was assessed one at a time.

Psychometric Analysis

The score sheets and candidate feedback forms were initially reviewed by Nelly. Questions that had candidate comments or were not answered correctly by the candidates were flagged for key validation and sent to the psychometric consultant for review.

Answer Key validation for the Numerical Assessment

Key validation is a comprehensive review of flagged questions by an SME panel. In the numerical assessment, 11 questions were flagged either due to candidate comments and/or none of the candidates answered the question correctly. As an additional measure, the remaining questions were also reviewed by the SMEs. The flagged questions were sent to two SMEs – one of them was the lead for the numerical assessment, who was involved with the development, standard setting, and scoring of the assessment, and an SME who was entirely new to the process. Both SMEs were practicing Approved Professionals in numerical standards. The results of the review process are described below. After carefully reviewing the flagged questions, the SMEs decided to remove two questions from the numerical-based assessment. The SMEs agreed that the candidate comments regarding these questions were valid and that both of these questions were flawed. The committee decided to remove 2 questions (worth 0.5 points each) from the exam. CSAP undertook the key validation and item review process with no additional assistance from the psychometrician. Upon removal of these questions the cut score was re-calculated and reviewed by the committee.

Answer Key validation for the Risk-based standards assessment

Eleven questions were flagged either due to candidate comments and/or none of the candidates answered the question correctly. As an additional measure, the remaining questions were also reviewed by the SMEs. The flagged questions were sent to two SMEs – one of them was the lead for the risk-based assessment, who was involved with the development, standard setting, and scoring of the assessment, and an SME who was entirely new to the process. Both SMEs were practicing Approved Professionals in risk-based standards. The results of the review process are described in the table below. Upon removal of these questions the cut score was re-calculated and reviewed by the committee. The committee decided to remove 7 questions from the exam. CSAP undertook the key validation and item review process with no additional assistance from the psychometrician.

Scoring Reliability

After key validation, the score sheets were sent to the psychometric consultant for analysis. Since a candidate was scored by two independent examiners, the psychometrician conducted inter-rater reliability analyses to examine the degree of agreement between the examiners when they scored the same candidate. The inter-rater reliability indices were satisfactory, indicating a high level of agreement between examiners.
The criteria for flagging a score discrepancy were if both examiners did not assign the exact same score to a candidate's response to a given question.

Numerical standards Oral Assessment

For each candidate, the scores provided by the two examiners needed to match completely. A total of 7 questions were flagged due to a score discrepancy.

Five candidates completed the Numerical exam. For the 1st candidate, 1 question was flagged due to scoring discrepancies. For the 2nd candidate, 6 questions were flagged. For the 3rd candidate, 2 questions were flagged. For the 4th candidate and 5th candidate, 3 questions were flagged, respectively. A scoring discrepancy was observed in 13 questions.

Score discrepancies were forwarded to a third examiner (a Numerical Standards Approved Professional who had not served as an examiner during the assessment, but who had completed the examiner training provided by the psychometric consultant). This third examiner listened to the audio recordings of the candidates for which there were score discrepancies in order to help them assign their scores. The scores provided by the third examiner were selected as the corrected scores for cases in which there were score discrepancies.

The table below summarizes the assessment statistics for the 2023 numerical standards oral assessment.

Table 23: *Final statistics for the 2023 numerical standards oral assessment*

	Statistic
Number of Examinees	5
Number of Scenarios	13
Number of scenario-based questions	50
Number of total marks	37
Raw Score Mean	26.3
Raw Score Standard Deviation	5.15
Raw Score Minimum	21.5
Raw Score Maximum	35.0
Reliability (Cronbach's Alpha) *	0.90
Raw Score Standard Error of Measurement (SEM)	0.13

***Please note:** Due to the very low sample size this statistic may not be accurate. Recommendations for the minimum number of examinees required to produce stable reliability results vary from 30 to 300, more than the number of examinees who sat this exam.

Risk-based standards Oral Assessment

For each candidate, the scores provided by the two examiners needed to match completely. A total of 23 questions were flagged due to a score discrepancy. Three candidates completed the Risk exam. For the 1st candidate, 11 questions were flagged due to scoring discrepancies. For the 2nd candidate, 19 questions were flagged. For the 3rd candidate, 14 questions were flagged. A scoring discrepancy was observed in 5 questions.

The table below summarizes the assessment statistics for the 2023 risk-based standards oral assessment.

Table 24: *Final statistics for the 2023 risk-based standards*

	Statistic
Number of Examinees	3
Number of Scenarios	16
Number of total marks	34
Number of scenario-based questions	35
Raw Score Mean	22.62
Raw Score Standard Deviation	1.89
Raw Score Minimum	20.60
Raw Score Maximum	25.15
Reliability (Cronbach's Alpha)*	0.34
Raw Score Standard Error of Measurement (SEM)	0.31

***Please note:** Due to the very low sample size this statistic may not be accurate. Recommendations for the minimum number of examinees required to produce stable reliability results vary from 30 to 300, more than the number of examinees who sat this exam.

Bibliography

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

Anderson, L., & Krathwohl, D. (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Kratwohl, D. R. (1956). *Taxonomy of educational objectives.* New York: Longmans Green.

Cizek, G.J. (2011). *Setting Performance Standards: Foundations, Methods, and Innovations*. New York, NY: Routledge.

Clauser, A.L., & Raymond, M. (2017). Specifying the content of credentialing examinations. In: Davis-Becker, S. & Buckendahl, C. (Eds.). *Testing in the Professions: Credentialing Policies and Practice*. New York, NY: Routledge. (pp. 64-84).

CSAP (n.d.). *About us*. https://csapsociety.bc.ca/about/

De Champlain, A. (2019). Standard Setting Methods in Medical Education: High Stakes Assessment. In: Swanwick, T., Forrest, K., & O'Brien, B. C. *Understanding Medical Education: Evidence, Theory, and Practice*. Oxford, U.K.: Wiley Blackwell. (pp. 347-359).

Davis-Becker, S. & Muckle, T.M. (2017). Test Design: Laying out the Roadmap. In: Davis-Becker, S. & Buckendahl, C. (Eds.). *Testing in the Professions: Credentialing Policies and Practice*. New York, NY: Routledge. (pp. 41-63).

Educational Testing Service (ETS). (2023). *Technical Manual for the PRAXIS Tests and Related Assessments.* Princeton, NJ.: ETS.

Haladyna, T.M. & Rodriguez, M.C. (2013). *Developing and Validating Test Items*. New York, NY: Taylor & Francis.

Hambleton, R. K., & Plake, B. S. (1995). Using an extended Angoff procedure to set standards on complex performance assessments. Applied Measurement in Education, 8, 41–55.

Haladyna TM, Shindoll RR. (1989). Item Shells: A Method for Writing Effective Multiple-Choice Test Items. *Evaluation & the Health Professions, 12,* 97-106.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1*(1), 3–62.

Parshall, C.G. & Brunner, B. (2017). Content Development and Review. In: Davis-Becker, S. & Buckendahl, C. (Eds.). *Testing in the Professions: Credentialing Policies and Practice*. New York, NY: Routledge. (pp. 41-63).

Riconscente, M.M., Mislevy, R.J., & Corrigan, S. (2015). Evidence-Centered Design. In: *Handbook of Test Development*. New York, NY: Routledge. (pp.).

Stone, Beltyukova, S. & Fox, C.M. (2008). Objective Standard Setting for Judge-Mediated Examinations. *International Journal of Testing, 8*, 180–196

Swanson, D.B., Norman, G.R., & Linn, R.L. (1995). Performance-Based Assessment: Lessons From the Health Professions. *Educational Researcher, 24*, 5-35.

Tekian, A. & Yudkowsky, R. (2009). Oral Assessments. In: Downing, S.& Yudkowsky, R. (Eds.). *Testing in Health Professions Education*. New York, NY: Routledge. (pp. 119-148).

Way, W.D. & Gialluca, A. (2017). In: Davis-Becker, S. & Buckendahl, C. (Eds.). Estimating, Interpreting, and Maintaining the Meaning of Test Scores. Testing in the Professions: Credentialing Policies and Practice. New York, NY: Routledge. (pp. 105-122).

Yudkowsky, R., Downing, S., & Tekian, A. (2009).Standard Setting. In: Downing, S.& Yudkowsky, R. (Eds.). Testing in Health Professions Education. New York, NY: Routledge. (pp. 119-148).

Zieky, M. (2014). An introduction to the use of evidence-centered design in test development. Psicología Educativa, 20, 79-87.

Appendix A: Practice Analysis Survey for Numerical standards

CSAP Numerical Standards

Practice Analysis survey

Instructions:

This survey is being conducted by the CSAP Society of B.C. The survey is part of a comprehensive study of the practice of numerical APs to inform the design and development of a valid, reliable and fair oral assessment of numerical standards. Please complete and return this form as soon as possible. The results of this survey will be aggregated to protect confidentiality.

Demographic information

Are you a voting member of the CSAP Society? Yes No

What year did you obtain CSAP certification? _____

How many years of CSAP experience do you have? _____

What is your CSAP specialty? _____

What is the name of the company you work for? _____

How long have you worked for this company? _____

Job location (city, province) _____

What is your job title? _____

What is your educational level? High School Bachelor's Degree Masters Degree or higher

What subject did you major in? _____

Optional:

Are you Indigenous or have any Indigenous heritage? Yes No

What is your gender identity? Female Male Non-Binary

What is your race/ethnicity? _____

What is your age? 18-30 31-40 41-50 51-60 60+

Is English your first language? Yes No

If not, what is your first language? _____

How many years have you been living and working in Canada? _____

Have you worked as a CSAP outside of Canada, if so, for how long? _____

Statement of Purpose for the Numerical Standards Oral Assessment

Instructions: Please review the statement of purpose below and provide feedback. Do you agree with the purpose of an oral assessment for numerical standards? Is there anything missing in this statement? Is there anything that is incorrect or irrelevant?

The purpose of the numerical standards oral assessment is to determine whether an examinee has sufficient technical knowledge in the assessment of contaminated sites, and the ability to make sound technical interpretations and decisions within the Contaminated Sites Regulation framework to conduct work as a Numerical Standards Approved Professional, including the submission of applications to the BC Ministry of Environment and Climate Change Strategy (ENV). Eligible work is defined under ENV Protocol 6.

Feedback:

Weighting of Oral Assessment Tasks

Instructions: The Numerical Standards Oral Assessment will be testing the following tasks via scenario-based questions. For each task, please provide a percent that indicates how much it should be weighted on the oral assessment. Tasks that have higher percentages will indicate greater weighting, which would mean these are more important tasks than tasks that have lower percentages. Please ensure your percentages add up to 100%.

_____Historical and Visual Site Information

e) APEC and PCOC: Identify all applicable potential APEC and PCOC based on review of existing information from various sources and based on assessment of site conditions observed during a site reconnaissance.

_____Assessment of Affected Media and Migration Pathway

i) Soil: Interpret site geology and soil stratigraphy.
j) Hydrogeology: Assess groundwater flow and contaminant transport (dissolved and Non-aqueous phase liquids - NAPL).
k) Surface hydrology: Interpret significance of precipitation on a contaminated site in terms of contaminant transport (surface water, groundwater, soil and sediment).
l) Sediment: Interpret sediment characteristics and its significance for contaminant distribution and release.
m) Soil vapour: Understand soil vapour concentrations and migration.
n) Air: Understand impact on indoor and outdoor air quality by dust and vapours from site contamination.
o) Biota: Understand significance of food-chain transfers and the significance of observations such as stressed vegetation and effects on aquatic life.

_____Contaminant Characteristics

d) Chemistry and biochemistry: Interpret physical, chemical and biological properties of contaminants and their significance on fate, transport, treatment and relative human health and ecological risks.
e) Chemical composition of mixtures: Understand the significance of chemical compositions of common types of contamination substances including but not limited to: fuels, lubricants, solvents, paints, wood preservatives, coal tar, metal plating, and landfill leachate.
f) Sources of Contamination: Be familiar with common residential, commercial and industrial activities that may result in site contamination including but not limited to: Fuel storage and handling, metal fabrication, wood preservation, solvent cleaning, coal gasification, and landfilling.

_____Investigation Approach and Methods

a) Sampling rationale: Interpret available information to develop a defensible sampling rationale that will satisfy the investigation objectives.

b) Sampling plans: Assess sampling plans to determine whether they are consistent with the investigation objectives and sampling rationale.

c) Sampling techniques: Understand the significance of the use of proper equipment and methods for sampling of soil, sediment, groundwater, surface water and soil vapour.
d) Field observations and records: Assess field records in terms of adequacy for data interpretation included but not limited to: Borehole logs, well installation details, visual/olfactory signs of contamination, sampling details, etc.
e) Laboratory testing methods: Understand applicability and limitations of common laboratory sampling methods including but not limited to: Gas chromatography, gas chromatography/mass spectroscopy, infrared spectroscopy, petroleum analytical methods (e.g., LEPH/HEPH vs. EPH).
f) Field screening techniques: Understand applicability and limitations of common laboratory sampling methods including but not limited to: soil vapour headspace, immunoassay, colorimetric, pH/conductivity/temperature, X-ray fluorescence.
g) QA/QC practices: Assess field and laboratory work in terms of acceptable QA/QC methods and interpretation.

_____Data Synthesis and Interpretation

a) Data integration and presentation: Assess the investigation data in terms of adequate presentation in tables and figures.
b) Adequacy of testing: Review sampling programs to assess the adequacy of the testing performed (number, type and location of samples).
c) Nature and extent of contamination: Assess APEC and AEC: number, types, characteristics, PCOC, delineation.
d) Nature and extent of migration pathways: Assess migration pathways: types, characteristics, preferential routes, relative importance.
e) Background conditions: Assess regional and local background conditions.

_____Risk Assessment Principles and Screening

a) Problem formulation: Identify/screen sources, exposure pathways, receptors
b) Acceptable risk: Carcinogens Non-carcinogens.
c) Exposure scenarios: Interpret current and future site uses.
d) High risk: Recognize imminent and high risk to human health and environment, and immediate risks to public welfare (e.g., explosion hazard, etc.).

_____Remediation Design, Implementation and Confirmation

a) Remediation techniques: Be familiar with common soil, sediment, groundwater, water and soil vapour remediation methods.
b) Remedial design: Understand technical, regulatory and cost aspects of common remediation methods, and be able to evaluate the selection of appropriate alternatives.
c) Remediation implementation: Understand health and safety standards, construction techniques/constraints, monitoring requirements, and requirements associated with off-site transport and disposal of contamination and record keeping.
d) Remediation Confirmation: Assess confirmatory sampling program and results in terms of adequacy to demonstrate the site meets the applicable requirements of a remediated site.

Please provide any comments about the content of the oral assessment. Is there anything missing from this list that should be tested in the assessment? Is there anything that is irrelevant or not required?

_____
_____
_____
_____

Knowledge Statements

Instructions: Please review the general knowledge statement below and provide feedback. Do you agree with this statement? Is there anything missing in this statement? Is there anything that is incorrect or irrelevant?

A successful candidate needs to understand and have the ability to apply combined aspects of soil science, environmental engineering, hydrogeology, environmental chemistry and basic risk assessment principles to review standards assessments and applicable documents to make recommendations based on application of numeric-based standards of the BC CSR. Candidates are also expected to have a general understanding of related areas.

Feedback:

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

Instructions: Please rate the degree to which each knowledge statement is needed to successfully accomplish the task of a numerical standards AP. How important is each knowledge in successfully performing the CSAP tasks?

1 = Not Important
2 = Marginally Important
3 = Moderately Important
4 = Important
5 = Critically Important

1. Know what to consider an APEC, and how to use historical data to get there.

   1 2 3 4 5

2. Know what PCOCs to assess for.

   1 2 3 4 5

3. How to properly assess a site (drilling, sampling, lab analysis)

   1 2 3 4 5

4. Be able to source information from the CSR and Guidance documents quickly.

   1 2 3 4 5

5. Understand the process for submission of legal instruments.

   1 2 3 4 5

6. Understand “101” level hydrogeology, chemistry, and biological “specialist” concepts.

   1 2 3 4 5

7. Understand the limitations of commonly used field equipment.

   1 2 3 4 5

8. Understand the general laboratory analytical processes conducted.

   1 2 3 4 5

9. Site Land use Information.

   1 2 3 4 5

10. Investigation methodology.

1 2 3 4 5

11. Contaminant identification/characterization.

1 2 3 4 5

12. Contaminant fate and transport.

1 2 3 4 5

13. Contaminant exposure pathways and risks to human health and environment.

1 2 3 4 5

14. Data interpretation and QA/QC.

1 2 3 4 5

15. Remediation planning and design.

1 2 3 4 5

16. Regulatory requirements and standards.

1 2 3 4 5

17. Understand the regulations and how they are implemented technically.

1 2 3 4 5

18. Understand the investigation, remediation and characteristics of common contaminants.

1 2 3 4 5

19. In-depth understanding of all historical activities that may have happened at a pulp mill.

1 2 3 4 5

20. Know every possible chemical involved in the production of paper.

1 2 3 4 5

21. How to complete a pumping test.

1 2 3 4 5

22. Identify information from the CSR and Guidance by memory.

1 2 3 4 5

23. Have knowledge of unwritten BC ENV protocol.

1 2 3 4 5

24. Understand complex hydrogeology, chemistry, and biological “specialist” concepts.

1 2 3 4 5

25. Understand the inner workings of field equipment.

1 2 3 4 5

26. Understand the detailed methodologies used for laboratory analysis.

1 2 3 4 5

27. Technology know-how of everything.

1 2 3 4 5

28. Non regulated substances.

1 2 3 4 5

29. Scope outside of the Environmental Management Act, Contaminated Site Regulation, protocol, guidance.

1 2 3 4 5

30. Detailed and complicated mathematical calculations or modelling.

1 2 3 4 5

31. Understand every single technical aspect of investigation, remediation and contaminants.

1 2 3 4 5

32. Understanding all characteristics of all possible contaminants.

1 2 3 4 5

33. Know potential contaminants of concern associated with most common land usage.

1 2 3 4 5

34. Know how to determine what processes/land uses are APECs and their related PCOCs.

1 2 3 4 5

35. Determine groundwater flow direction using representative wells.

1 2 3 4 5

36. Conduct aquifer tests and determine hydrogeological conditions.

1 2 3 4 5

37. Interpret response test data.

1 2 3 4 5

38. Know when to go to an expert.

1 2 3 4 5

39. U understand preferential pathways.

1 2 3 4 5

40. Understand what a conductivity value means.

1 2 3 4 5

41. Understand the limitations of response test methods.

1 2 3 4 5

42. Identify the physical/chemical properties that affect fate and transport of PCOCs.

1 2 3 4 5

43. Find physical and chemical properties of contaminants.

1 2 3 4 5

44. Verify proper sampling methods, field screening, laboratory tests, and holding times.

1 2 3 4 5

45. Apply different interpretation methods.

1 2 3 4 5

46. Review and evaluate QA/QC data.

1 2 3 4 5

47. Develop a conceptual site model.

1 2 3 4 5

48. Run bench scale and pilot scale tests.

1 2 3 4 5

49. Identify data gaps and their significance.

1 2 3 4 5

50. Determine what standards apply to a site.

1 2 3 4 5

51. Use Stage 1 PSI conclusions to plan Stage 2 investigation.

1 2 3 4 5

52. Use Stage 2 PSI results to plan DSI program.

1 2 3 4 5

53. Develop a sampling plan.

1 2 3 4 5

54. Identify basic chemical properties.

1 2 3 4 5

55. Plan a remedial excavation program.

1 2 3 4 5

56. Know the ethical/disciplinary policies of the governing body and when to report.

1 2 3 4 5

57. Be comfortable with nuances of investigation rational and methods

1 2 3 4 5

58. Be able to find technical info on the ENV website.

1 2 3 4 5

59. Know current and emerging remediation technologies and their application for most contaminants.

1 2 3 4 5

60. Know physical/chemical properties and what they mean.

1 2 3 4 5

61. Understand health/safety standards and constraints.

1 2 3 4 5

62. Demonstrate mathematical competence.

1 2 3 4 5

63. Know most investigative and remediation limitations.

1 2 3 4 5

64. Be able to source stage 1 info sources.

1 2 3 4 5

65. Know standard investigative procedures and remedial approaches and who to ask.

1 2 3 4 5

66. Be aware of own limitations and limitations of reliance on data for their site.

1 2 3 4 5

67. Be able to back out of situations and ask others as necessary.

1 2 3 4 5

68. Recognize gray areas in terms of data interpretation.

1 2 3 4 5

69. Be competent in fundamental fields involved in profession, e.g., hydrogeology, chemistry.

1 2 3 4 5

70. Provide interpretations of investigation results which are usually correct and don't make fatal errors.

1 2 3 4 5

71. Be allowed to make some document, administrative or investigation errors but no fatal errors.

1 2 3 4 5

72. Have been exposed to most situations but may need to look up things.

1 2 3 4 5

73. Know to go to several sources.

1 2 3 4 5

74. Know when judgment is allowed.

1 2 3 4 5

75. Check technical references before performing services.

1 2 3 4 5

76. Review own work before submitting it.

1 2 3 4 5

Comments

Please provide any comments about the knowledge statements. Is there anything missing from this list? Is there anything that is irrelevant or not required?

Do you have any final comments that you would like to share with the CSAP Society regarding this survey or the development of the oral assessment?

Appendix B: Practice Analysis Survey for Risk-based standards

CSAP Risk-Based Standards

Practice Analysis survey

Instructions:

This survey is being conducted by the CSAP Society of B.C. The survey is part of a comprehensive study of the practice of Risk-based Standards APs to inform the design and development of a valid, reliable and fair oral assessment of risk-based standards. Please complete and return this form as soon as possible. The results of this survey will be aggregated to protect confidentiality.

Demographic information

Are you a voting member of the CSAP Society? Yes No

What year did you obtain CSAP certification? _____

How many years of CSAP experience do you have? _____

What is your CSAP specialty? _____

What is the name of the company you work for? _____

How long have you worked for this company? _____

Job location (city, province) _____

What is your job title? _____

What is your educational level? High School Bachelor's Degree Masters Degree or higher

What subject did you major in? _____

Optional:

Are you Indigenous or have any Indigenous heritage? Yes No

What is your gender identity? Female Male Non-Binary

What is your race/ethnicity? _____

What is your age? 18-30 31-40 41-50 51-60 60+

Is English your first language? Yes No

If not, what is your first language? _____

How many years have you been living and working in Canada? _____

Have you worked as a CSAP outside of Canada, if so, for how long? _____

Statement of Purpose for the Risk-Based Standards Oral Assessment

Instructions: Please review the statement of purpose below and provide feedback. Do you agree with the purpose of an oral assessment for numerical standards? Is there anything missing in this statement? Is there anything that is incorrect or irrelevant?

The purpose of the risk assessment oral assessment is to determine whether an examinee has sufficient technical knowledge in the assessment of contaminated sites, and the ability to make sound technical interpretations and decisions within the Contaminated Sites Regulation framework to conduct work as a Risk Assessment Standards Approved Professional including the submission of applications to the BC Ministry of Environment and Climate Change Strategy (ENV). Eligible work is defined under the ENV Protocol 6.

Feedback:

Weighting of Oral Assessment Tasks

Instructions: The Risk-Based Standards Oral Assessment will be testing the following tasks via scenario-based questions. For each task, please provide a percent that indicates how much it should be weighted on the oral assessment. Tasks that have higher percentages will indicate greater weighting, which would mean these are more important tasks than tasks that have lower percentages. Please ensure your percentages add up to 100%.

Ecological Risk Assessment

_____Problem Formulation

p) Risk Assessment Planning
q) Integration of Available Information
r) Identification of stressors
s) Potentially Exposed Receptors
    a. Complete and incomplete pathways
    b. Risk controls
t) Selecting Assessment and Measurement Endpoints
u) Conceptual Models
v) Data Gap Analysis
w) Sampling and Analysis Plan

_____Exposure Assessment

d) Characterization of Exposure
e) Evaluating Data and Models for Analysis
    a. Strengths and Limitations of Different Types of Data
    b. Literature Data – relevant species, study conditions
    c. Site Data/Observations - measurement and assessment endpoints; species diversity,
       richness, abundance
f) Measurement and/or Modeling Studies

_____Effects Assessment

f) Quantitative and Qualitative Site Observations
    a. Terrestrial Receptors
    b. Aquatic Receptors
g) Bioassays
    a. Field studies
    b. Laboratory toxicity tests
h) Toxicity Reference Values
    a. Selection
    b. Derivation
i) Ecosystem – context of scale relative to contaminated sites
j) Ecological Responses
    a. Stressor-Response Analysis
    b. Establishing Cause-and-Effect Relationships

c. Linking Measures of Effect to Assessment Endpoints

_____Risk Characterization

e) Quotient Method
f) Observation Method
g) Weight of Evidence
h) Reporting Risks

_____Uncertainty Analysis

d) Identifying Major Types of Uncertainty
e) Use of Uncertainty Factors
f) Sensitivity Analysis

Human Health Risk Assessment

_____Problem Formulation

j) Data Collection
    a. Background Information Useful for Data Collection
    b. Review of Available Site Information
    c. Addressing Modeling Parameter Needs
    d. Preliminary Identification of Potential Human Exposure
    e. Strategy for Sample Collection
    f. QA/QC Measures
k) Data Evaluation
    a. Combining Data Available from Site Investigations
l) Evaluation of Analytical Methods
m) Evaluation of Quantitation Limits
n) Chemicals of Potential Concern
    a. Comparison of Samples with Criteria/Guidelines
    b. Comparison of Samples with Standards
o) Potentially Exposed Receptors
p) Potential Exposure Pathways
    a. Complete and incomplete pathways
    b. Risk controls
q) Conceptual Model
r) Data Gap Analysis

_____Exposure Assessment

f) Characterization of Exposure Setting
    a. Characterize Physical Setting
    b. Characterize Exposed Receptors
    c. Identification of Exposure Routes
    d. Identification of Reasonable Maximum Exposure
g) Quantification of Exposure: Determining Exposure Concentrations

a. Estimation of Chemical Intakes
b. Exposure Concentrations in Various Media
c. Combining Chemical Intakes Across Pathways

_____Toxicity Assessment

h) Types of Toxicological Information Considered in Toxicity Assessment
i) Toxicity Assessment for Noncarcinogenic Effects
j) Toxicity Assessment for Carcinogenic Effects
k) Identifying Appropriate Toxicity Values for Site Risk Assessment
l) Evaluating Chemicals for which no Regulatory Toxicity Values are Available

_____Risk Characterization

f) Quantifying Risks
    a. Risks for Individual Substances
    b. Risks for Multiple Substances
g) Combining Risks Across Exposure Pathways
h) Consideration of Site-Specific Human Studies
i) Risk Characterization Results
j) Risk controls

_____Uncertainty Analysis

d) Identifying Major Types of Uncertainty
e) Use of Uncertainty Factors
f) Sensitivity Analysis

Please provide any comments about the content of the oral assessment. Is there anything missing from this list that should be tested in the assessment? Is there anything that is irrelevant or not required?

_____
_____
_____
_____

Knowledge Statements

Instructions: Please review the general knowledge statement below and provide feedback. Do you agree with this statement? Is there anything missing in this statement? Is there anything that is incorrect or irrelevant?

A successful candidate needs to understand and have the ability to apply combined aspects of ecology, toxicology and environmental chemistry to review of human health and ecological risk assessments and applicable documents to make recommendations based on application of risk-based standards of the BC CSR. Candidates are also expected to have a general understanding of related areas such as for example, basic contaminant transport in various media.

Feedback:

Instructions: Please rate the degree to which each knowledge statement is needed to successfully accomplish the task of a risk-based standards AP. How important is each knowledge in successfully performing the CSAP tasks?

1 = Not Important
2 = Marginally Important
3 = Moderately Important
4 = Important
5 = Critically Important

77. Know what “arm’s length review” is.

    1 2 3 4 5

78. Be able to review risk assessment and identify where rationale is inadequate.

    1 2 3 4 5

79. Know key ministry guideline documents.

    1 2 3 4 5

80. Explain all decisions; provide rationale behind things.

    1 2 3 4 5

81. Explain development of federal standards and provincial guidelines; know the history.

    1 2 3 4 5

82. Make submissions independently under Protocol 6 (administrative work, flawless submissions).

    1 2 3 4 5

83. Define a chemical carcinogenic substance and how it differs in different areas.

    1 2 3 4 5

84. Describe contaminant chemistry and its influence on fate and transport.

    1 2 3 4 5

85. Explain fundamental toxicological principles (adsorption, transformation, target tissue effects, excretion).

    1 2 3 4 5

86. Explain fundamental ecological/biological concepts relating to quantitative methods used in ecological risk assessment.

1 2 3 4 5

87. Use fundamental dose equations (e.g., Health Canada, ORNL).

1 2 3 4 5

88. Explain the different types of human health toxicological reference values, i.e. RFD, tolerable concentration, slope factor, unit risk (i.e. what they represent).

1 2 3 4 5

89. Use and convert these values in order to calculate human health risks (HQ and ILCR).

1 2 3 4 5

90. Know the basic standard measures of toxicity to ecological receptors (e.g. LC50, EC20, IC20, NOEC, LOEC, NOAEL, LOAEL).

1 2 3 4 5

91. Describe the difference between an HQ and ILCR.

1 2 3 4 5

92. Characterize risk; calculate risk estimates.

1 2 3 4 5

93. Understand how exposure concentrations should be selected for different receptors.

1 2 3 4 5

94. Pick the right variable for an equation.

1 2 3 4 5

95. Pick a TRV from a source.

1 2 3 4 5

96. Know how that TRV was calculated, and how to leverage that information in rationale.

1 2 3 4 5

97. Determine if certain risk assessments are worthy of being submitted under Protocol 6.

1 2 3 4 5

98. Explain fundamental ecological/biological concepts relating to quantitative methods used in ecological risk assessment.

1 2 3 4 5

99. Use fundamental dose equations (e.g., Health Canada, ORNL).

1 2 3 4 5

100. Explain the different types of human health toxicological reference values, i.e. RFD, tolerable concentration, slope factor, unit risk (i.e. what they represent).

1 2 3 4 5

101. Able to apply regulatory requirements.

1 2 3 4 5

102. Determine if risk assessment is worthy of submission under protocol 6.

1 2 3 4 5

Comments

Please provide any comments about the knowledge statements. Is there anything missing from this list? Is there anything that is irrelevant or not required?

Do you have any final comments that you would like to share with the CSAP Society regarding this survey or the development of the oral assessment?

Appendix C: Summary of results of the Numerical standards practice analysis survey

Comments about the Statement of Purpose

- assessment "and remediation" of contaminated sites, technical "judgement", interpretation.
- Looks fine to me.
- Yes
- I don't agree with the use of an oral assessment as a stand-alone evaluation of CSAP candidates technical knowledge for the following reasons: 1. For those who have English as a second or third language, or have a speech impediment it puts them at a disadvantage. 2. Oral assessments put them on the spot in front of evaluators and don't provide time and space to consider the question thoroughly and without pressure. People who are not comfortable with public speaking or have higher levels of anxiety will be at a disadvantage. 3. Technical answers are extremely nuanced. As a previous content creator I can say that a lot of thought, vetting, review and discussion goes into the construction/structuring of each of the questions in the current format of the Numerical exam in order to construct a question which has only one answer, and the creation of answers of which only one is correct and the rest are incorrect. Under an oral assessment a candidate may provide several answers for one question based on variables that are not part of the question, and may not have been considered as part of the question construction. In other cases answers provided may be correct, partially correct, or correct but contain incorrect underlying statements. Accordingly, it will be extremely hard to provide a quantitative evaluation of the candidates response, leading to a qualitative evaluation of the candidates response, which is then subjective based on the level of competency of the oral evaluator(s). This subjectivity/qualitative assessment was highlighted as an big issue in the evaluation of the Roster/CSAP exam when it was previously provided in an "essay format" setting. This lead us to seeking professional testing advice, which was why the format was changed to multiple choice. I wrote the exam under both formats and can say that the change was a significantly positive one. 4. All previously developed content for the Numerical exam would have to be significantly reworked for an oral format to the point where we would essentially be starting from scratch. Huge time and budget commitment would be required to make this shift.

Weighting results

	Minimum (%)	Maximum (%)	Average (%)	Standard Deviation
Historical and Visual Site Information	5	15	9.5	3.708099
Assessment of Affected Media and Migration Pathway	17	20	18.8	1.643168
Contaminant Characteristics	10	20	15.4	3.646917
Investigation Approach and Methods	15	25	19	3.807887
Data Synthesis and Interpretation	10	20	15.4	3.646917
Risk Assessment Principles and Screening	5	12	7.4	3.361547
Remediation Design, Implementation and Confirmation	12.5	17	14.5	1.802776

For reference, I have included the current weighting that the written exam was based on.

Content Area	**2015 Syllabus Weight %**	**New Weights**
Historical and Visual Site Information	5.0%	9.5%
Assessment of Affected Media and Migration Pathway	20.0%	18.8%
Contaminant Characteristics	17.0%	15.4%
Investigation Approach and Methods	17.0%	19.0%
Data Synthesis and Interpretation	17.0%	15.4%
Risk Assessment Principles and Screening	7.0%	7.4%
Remediation Design, Implementation and Confirmation	17.0%	14.5%
TOTAL	100%	100%

Comments about the weighting

- would be good to add a question or two on screening level risk assessment
- Add "depositional environment" to "Assessment of Affected Media..." category (subcategory "a"). Add "hydraulic gradient and direction" to "Assessment of Affected Media..." category (subcategory "b"). i am not sure what "Understand significance of food chain transfers..." means? How is this relevant? Add "wildlife" to "Assessment of Affected Media..." category (subcategory "g"): "...stressed vegetation and effects on wildlife and aquatic life". Add a subcategory "f" under the "Data Synthesis..." category with the following info: "f) concentration trend analysis - understand results wrt stable, decreasing, increasing and indeterminant trends and their significance to overall site conditions". Comment for "Remediation Design..." category (subcategory "d"): "perhaps understanding trend analysis and results should be in this section?"
- The tasks structuring hasn't really changed since the inception of the CSAP/Roster exams in the late 1990s. A lot has changed since then in the overall understanding and competency of contaminated sites practitioners, and how contaminated sites are

approached. The task structuring should really be updated/reorganized to reflect a more current approach to viewing the thought process used to evaluate contaminated sites problems based on a Conceptual Site Model (CSM) approach, i.e. evaluation of the Source, Transport, and Exposure components of the CSM with the addition of remediation (which includes Risk Assessment). Relevant weightings for those components (in my mind) would be 30%, 30%, 20% and 20% respectively. The sub-sections under these new main sections would have to be reviewed, but would contain most of the main sub-sections as currently noted, or in concept. The issue with the current Task Structure becomes quite apparent when setting the exams as some tasks contain multiple variations of what is essentially the same concept, and writing new questions is quite difficult as the task is quite narrow in its requirements. A good example of this is data synthesis and interpretation. Sub-task d) of this task should really be under the Assessment of Migration task. The remaining sub-tasks really duplicate concepts which are commonly tested within the Regulatory exam. They are also of limited scope, and as such it is increasingly difficult to create more content for this task.

Comments about the General Knowledge Statement

- add: "contaminant fate and transport" to the sentence: "….environmental chemistry, contaminant fate and transport, and basic risk assessment principles,…"
- Last sentence is very vague. Is it necessary? If so, what are the related areas? Aren't they the ones listed in the first sentence?
- Yes
- The Environmental (Contaminated sites) field is a multidisciplined field of work/study. While not every project/problem requires the involvement of SMEs in all of the sub-disciplines, every project requires a generalist with a knowledge of all of the sub-disciplines and how they integrate into the project. While a numerical CSAP will likely be a SME in one or more of the sub-disciplines, the requirement is that they be a high-level generalist, capable of reviewing the work of other generalists which may not have all the knowledge of all sub-disciplines. The CSAP should be fully capable of determining when input of SMEs is required and be able to understand the conclusions and recommendations of the SMEs (but not necessarily the underlying technical detail), and how they apply to the project. Per the list of sub-disciplines provided in the above statement, I would change that to read: "....combined aspects of basic soil science, hydrogeology, hydrology, bio-chemistry, geo-chemistry, biology, toxicology, and common industrial processes..". As currently listed, env. chemistry, hydrogeology, and soil science are very specific areas of study, whereas environmental engineering and risk assessment are broad areas of study encompassing env. chemistry, hydrogeology, and soil science.

Evaluation of Knowledge Statements

(Sorted by Average rating from highest to lowest)

	Minimum	Maximum	Average	Standard Deviation
1. Know what to consider an APEC, and how to use historical data to get there.	5	5	5	0
2. Know what PCOCs to assess for.	5	5	5	0
5. Understand the process for submission of legal instruments.	5	5	5	0
16. Regulatory requirements and standards.	5	5	5	0
18. Understand the investigation, remediation and characteristics of common contaminants.	5	5	5	0
32. Know potential contaminants of concern associated with most common land usage.	5	5	5	0
33. Know how to determine what processes/land uses are APECs and their related PCOCs.	5	5	5	0
34. Determine groundwater flow direction using representative wells.	5	5	5	0
37. Know when to go to an expert.	5	5	5	0
43. Verify proper sampling methods, field screening, laboratory tests, and holding times.	5	5	5	0
46. Develop a conceptual site model.	5	5	5	0
48. Identify data gaps and their significance.	5	5	5	0
49. Determine what standards apply to a site.	5	5	5	0
53. Identify basic chemical properties.	5	5	5	0
65. Be aware of own limitations and limitations of reliance on data for their site.	5	5	5	0
66. Be able to back out of situations and know who to ask if necessary.	5	5	5	0
72. Know when judgment is allowed.	5	5	5	0
74. Review own work before submitting it.	5	5	5	0
3. How to properly assess a site (drilling, sampling, lab analysis)	4	5	4.75	0.5
4. Be able to source information from the CSR and Guidance documents.	4	5	4.75	0.5

6. Understand "101" level hydrogeology, chemistry, and biological "specialist" concepts.	4	5	4.75	0.5
9. Historical, current and potential future site land use Information.	4	5	4.75	0.5
11. Contaminant identification/characterization.	4	5	4.75	0.5
17. Understand the regulations and how they are implemented technically.	4	5	4.75	0.5
39. Understand what a hydraulic conductivity value means.	4	5	4.75	0.5
42. Find physical and chemical properties of contaminants.	4	5	4.75	0.5
50. Use Stage 1 PSI conclusions to plan a Stage 2 investigation.	4	5	4.75	0.5
51. Use Stage 2 PSI results to plan a DSI program.	4	5	4.75	0.5
52. Develop a sampling plan.	4	5	4.75	0.5
64. Know standard investigative procedures and remedial approaches.	4	5	4.75	0.5
67. Recognize gray areas in terms of data interpretation.	4	5	4.75	0.5
70. Have been exposed to most situations but may need to look up things.	4	5	4.75	0.5
7. Understand the limitations of commonly used field equipment.	4	5	4.5	0.57735
10. Investigation methodology.	4	5	4.5	0.57735
12. Contaminant fate and transport.	4	5	4.5	0.57735
19. In-depth understanding of all historical activities that may have happened at a site.	4	5	4.5	0.57735
38. Understand preferential pathways.	4	5	4.5	0.57735
56. Be comfortable with nuances of investigation rational and methods	3	5	4.5	1
59. Know physical/chemical properties and what they mean.	3	5	4.5	1
62. Know most investigative and remediation limitations.	4	5	4.5	0.57735
69. Be allowed to make some document, administrative or investigation errors but no fatal errors.	4	5	4.5	0.57735
71. Know to go to several sources.	4	5	4.5	0.57735

73. Check technical references before performing services.	4	5	4.5	0.57735
14. Data interpretation and QA/QC.	3	5	4.25	0.957427
41. Identify the physical/chemical properties that affect fate and transport of PCOCs.	4	5	4.25	0.5
45. Review and evaluate QA/QC data.	3	5	4.25	0.957427
57. Be able to find technical information on the ENV website.	3	5	4.25	0.957427
68. Be competent in fundamental fields involved in profession, e.g., hydrogeology, chemistry.	4	5	4.25	0.5
8. Understand the general laboratory analytical processes conducted.	3	5	4	0.816497
55. Know the ethical/disciplinary policies of the governing body and when to report.	2	5	4	1.414214
13. Contaminant exposure pathways and risks to human health and environment.	3	4	3.75	0.5
15. Remediation planning and design.	3	4	3.75	0.5
40. Understand the limitations of response test methods.	3	4	3.75	0.5
44. Apply different methods of interpretation.	3	4	3.75	0.5
54. Plan a remedial excavation program.	3	5	3.75	0.957427
63. Be able to source Stage 1 PSI information sources.	3	4	3.75	0.5
31. Understanding all characteristics of all possible contaminants.	2	5	3.5	1.732051
58. Know current and emerging remediation technologies and their application for most contaminants.	3	5	3.5	1
60. Understand health/safety standards and constraints.	2	5	3.5	1.290994
35. Conduct aquifer tests and determine hydrogeological conditions.	3	4	3.25	0.5
61. Demonstrate mathematical competence.	2	5	3.25	1.5
28. Scope outside of the Environmental Management Act,	2	4	3	0.816497

Contaminated Site Regulation, protocol, guidance.				
23. Understand complex hydrogeology, chemistry, and biological “specialist” concepts.	2	3	2.5	0.57735
25. Understand the detailed methodologies used for laboratory analysis.	2	3	2.5	0.57735
27. Non regulated substances.	1	4	2.5	1.290994
30. Understand every single technical aspect of investigation, remediation and contaminants.	1	4	2.5	1.290994
36. Interpret response test data.	1	3	2.5	1
22. Have knowledge of unwritten BC ENV protocol.	1	3	2.25	0.957427
20. How to complete a pumping test.	2	2	2	0
29. Detailed and complicated mathematical calculations or modelling.	1	3	2	0.816497
24. Understand the inner workings of field equipment.	1	3	1.666667	1.154701
21. Identify information from the CSR and Guidance by memory.	1	2	1.5	0.57735
47. Run bench scale and pilot scale tests.	1	3	1.5	1
26. Technology know-how of everything.	1	2	1.333333	0.57735

Comments about the knowledge statements

- 8. Understand the general laboratory analytical processes conducted. (change "processes" to "testing"). 26. Technology know-how of everything. (?? what is everything?). 28. Scope outside of the Environmental Management Act, Contaminated Site Regulation, protocol, guidance. (scope of what? other jurisdictions/regulations?). 36. Interpret response test data. (different from "reduce the data" which would be a 1 or 2.).
- 4. Be able to source information from the CSR and Guidance documents quickly. (It is not the speed of sourcing the information but the relevance and accuracy that is important). 9. Site Land use Information. (Do you mean historical as well as current and future activities?). 19. In-depth understanding of all historical activities that may have happened at a pulp mill. (Why pulp mill? This is also something that the candidate can research, and should not need to know for this exam). 31. Understand every single technical aspect of investigation, remediation and contaminants. (Can go to expert or research self.). 40. Understand what a conductivity value means. (Hydraulic conductivity). 45. Apply different interpretation methods. (Not sure what you are looking for here. I don't interpretation is a method.). 70. Provide interpretations of investigation

results which are usually correct and don't make fatal errors. (These kind of questions speaks to performance of own work and not knowledge to demonstrate for a successful candidate. Indirectly, if the candidate meet the knowledge above, they would not make fatal errors.).

- The list is quite long and unstructured. Not really sure what you are trying to achieve with the knowledge statements so it is hard to intelligibly add or comment to this list. There are many statements provided at many different levels of detail.

Final comments

- Again, I'm not sure why we are moving to an oral assessment. It does seem like we are moving backward within this approach, discarding an approach which was recommended by a 3rd party exam professionals, and to which CSAP has dedicated a great deal of time and money to develop. I'm not aware of any significant complaints with the current format. My recommendation would be to stick with the current format and where necessary provide the changes necessary within that framework to make improvements.
- I have a concern that the oral assessment may be somewhat subjective.

Appendix D: Summary of results of the Risk-based standards practice analysis survey

Comments about the Statement of Purpose

- could add the current version of Protocol 6
- agree
- Definition is reasonable and does not contain irrelevant or incorrect information. I would amplify on the idea that the goal is to determine if the examinee can make sound technical decisions. I think that means our focus is to briefly confirm that the projects submitted as experience provide sufficient breadth of practice, but spend most of the effort on determining that the examinee understands a) how to make informed decisions in light of uncertainty, but b) what are the limits of the allowable discretion in decision making as prescribed by the regulation.
- Seems reasonable but I'm not clear on whether the oral assessment is a component of the exam or only format of the exam. If the former, then some additional purpose might be appropriate, e.g., " …. sufficient technical knowledge to verbally articulate technical knowledge in the assessment…"). If it is being considered as the only format, then this statement is fine as written.
- I generally agree; however it may be difficult to test some of these in an oral assessment but I would be interested in learning the type of questions that will be asked and then more important how they are marked. Please do not get me wrong - I'm very supportive of this undertaken!
- change "assessment of contaminated sites" to "risk assessment of contaminated sites". very broad purpose

Weighting results

	Minimum (%)	Maximum (%)	Average (%)	Standard Deviation
Ecological Risk Assessment	50	50	50	0
Problem Formulation	15	30	20	8.660254
Exposure Assessment	10	25	15.83333	8.036376
Effects Assessment	10	25	15.83333	8.036376
Risk Characterization	7.5	20	14.16667	6.291529
Uncertainty Analysis	0	2.5	1.25	1.767767
Human Health Risk Assessment	50	50	50	0
Problem Formulation	10	25	15.83333	8.036376
Exposure Assessment	15	25	18.33333	5.773503
Toxicity Assessment	10	25	15	8.660254
Risk Characterization	10	25	16.66667	7.637626
Uncertainty Analysis	0	2.5	1.25	1.767767

The weighting results of only 3 participants were used since the weightings of the other participants did not total 100%. For reference, I have included the current weighting that the written exam was based on:

Content Area	**2015 Syllabus Weight %**	**Weighting by participant 1**	**Weighting by participant 2**	**Weighting by participant 3**
Ecological Risk Assessment	**50%**	**50%**	**50%**	**50%**
Problem Formulation	15.0%	30.0%	15.0%	15.0%
Exposure Assessment	12.5%	25.0%	12.5%	10.0%
Effects Assessment	12.5%	25.0%	12.5%	10.0%
Risk Characterization	7.5%	20.0%	7.5%	15.0%
Uncertainty Analysis	2.5%	0	2.5%	0
Human Health Risk Assessment	**50%**	**50%**	**50%**	**50%**
Problem Formulation	12.5%	25.0%	12.5	10.0%
Exposure Assessment	15.0%	25.0%	15.0%	15.0%
Toxicity Assessment	10.0%	25.0%	10.0%	10.0%
Risk Characterization	10.0%	25.0%	10.0%	15.0%
Uncertainty Analysis	2.5%	0	2.5%	0
TOTAL	**100%**	**100%**	**100%**	**100%**

Comments about the weighting

- There are items in Protocol 1 that are not captured here and will not be included in the Regulatory exam as they are too RA specific. For example, 1) Protocol 28 as the default source for TRVs and human receptor characteristics, 2) Detailed risk assessment report

submission requirements: requirements for report completeness, errors and omissions (table 2). Items such as the above need to be captured somewhere in the oral assessment.

- I think it's critical that scenario-based questions are not geared to probing for the "right answer". I'm also not convinced that the AP needs to be examined on all aspects of risk assessment - each candidate has already passed the hurdle of showing that they have submitted multiple risk assessments and have 10+ years of experience. I think the questions need to be focused on understanding how the candidate is using risk assessment guidance in their individual practice - there is rarely one "right" way to solve an issue.
- In general, it would seem appropriate to limit the number of verbal responses requiring calculations and focus more on reasoning and understanding of concepts.
- I put 0 for uncertainty because it is just not very testable in any meaningful manner (uncertainty analysis is still an important part of risk assessment; however, it is just not very testable)
- missing is the performance verification plan and risk management recommendations. These topics were not included historically and are considered a gap in the testing for approved professionals.

Comments about the General Knowledge Statement

- A successful candidate needs to understand and have the ability to apply combined aspects of environmental chemistry, toxicology, and basic contaminants transport in various media to the review of human health and ecological risk assessments and other applicable documents (e.g. performance verification plans) to make recommendations based on application of the risk-based standards of the BC CSR. Candidates are also expected to have a general understanding of related areas such as for example, site investigation (e.g. sampling) of various media.
- I think the statement is fine. The list of related areas is much more extensive, but the premise that risk assessment requires the integration of chemistry, toxicology and ecology is correct.
- "…to review human health and ecological risk..". This statement is fine as written but I'm not clear on how it would change for an oral assessment format.
- I generally agree with this statement. Where it gets a bit tricky is how much of related areas is required to be known; however, I think it is still fair statement as an AP needs to make decisions using this information. Nevertheless, if exam candidates were given specific ideas of the type of knowledge areas ahead of time, it would be a win-win (i.e., they could study and learn these and that will make them even better APs)
- Candidates need to have a basic understanding of site characterization of soil, sediment, groundwater, surface water, and vapour and the ability to understand if there are data gaps that will limit the risk assessment and create uncertainty for the risk conclusions.

Evaluation of Knowledge Statements

(Sorted by Average rating from highest to lowest)

Knowledge Statement	Minimum	Maximum	Average	Standard Deviation
1. Know what “arm’s length review” is.	5	5	5	0
15. Describe the difference between an HQ and ILCR.	5	5	5	0
2. Be able to review risk assessment and identify where rationale is inadequate.	5	5	5	0
21. Determine if certain risk assessments are worthy of being submitted under Protocol 6.	5	5	5	0
26. Determine if risk assessment is worthy of submission under protocol	5	5	5	0
3. Know key ministry guideline documents.	5	5	5	0
11. Use fundamental dose equations (e.g., Health Canada, ORNL).	4	5	4.8	0.447214
16. Characterize risk; calculate risk estimates.	4	5	4.8	0.447214
23. Use fundamental dose equations (e.g., Health Canada, ORNL).	4	5	4.75	0.5
13. Use and convert these values in order to calculate human health risks (HQ and ILCR).	3	5	4.6	0.894427
14. Know the basic standard measures of toxicity to ecological receptors (e.g. LC50, EC20, IC20, NOEC, LOEC, NOAEL, LOAEL).	3	5	4.6	0.894427
17. Understand how exposure concentrations should be selected for different receptors.	4	5	4.6	0.547723
18. Pick the right variable for an equation.	4	5	4.6	0.547723
25. Able to apply regulatory requirements.	4	5	4.6	0.547723
22. Explain fundamental ecological/biological concepts relating to quantitative methods used in ecological risk assessment.	4	5	4.5	0.57735
24. Explain the different types of human health toxicological reference values, i.e. RFD, tolerable concentration, slope factor, unit risk (i.e. what they represent).	3	5	4.5	1

12. Explain the different types of human health toxicological reference values, i.e. RFD, tolerable concentration, slope factor, unit risk (i.e. what they represent).	3	5	4.4	0.894427
10. Explain fundamental ecological/biological concepts relating to quantitative methods used in ecological risk assessment.	3	5	4.2	0.83666
6. Make submissions independently under Protocol 6 (administrative work, flawless submissions).	3	5	4.2	0.83666
19. Pick a TRV from a source.	2	5	4	1.224745
4. Explain all decisions; provide rationale behind things.	3	5	4	0.816497
20. Know how that TRV was calculated, and how to leverage that information in rationale.	2	5	3.8	1.095445
7. Define a chemical carcinogenic substance and how it differs in different areas.	3	5	3.8	0.83666
9. Explain fundamental toxicological principles (adsorption, transformation, target tissue effects, excretion).	3	5	3.8	0.83666
5. Explain development of federal standards and provincial guidelines; know the history.	2	5	3.6	1.140175
8. Describe contaminant chemistry and its influence on fate and transport.	3	4	3.2	0.447214

Comments about the knowledge statements

- understand when risk estimates should be added across more than one COPC, exposure pathways, and/or receptor.
- The knowledge statements are fine as is but do not provide additional context or rationale regarding the oral assessment component if that was the intent.

Final comments

- the section on weighting is based on the RA exam syllabus, which has differing levels of detail depending on the category/sub-category. It may be necessary to revise or prepare a new RA exam syllabus for the oral assessment process. The current version does not appear to be posted on the CSAP website any longer.

- I'm not clear on the rationale for using oral assessment in the exam process - what advantage does it have over the written/computer based exam and is another skill being sought from the examinees? If some competency in verbal articulation is being sought then the scoring will need to be carefully developed for fairness for all potential candidate types. Candidates who have not had this type of assessment before or candidates with English as a second language, for example, may have more difficulty with the format and would score less even though they have similar knowledge levels. I also don't believe that testing for competency in verbal articulation of knowledge is a needed in risk-based CSAP exams. In practice, most candidates are consultants and have some level of skill in this area and most communications are in writing.
- I think one of the more important aspects is ensuring that a person really has the 10,000 hours of risk assessment experience that is required to be a risk AP. And in tallying these 10,000 hours, making sure it is risk assessment and not site investigation. I also think that there may an opportunity to include questions on instruments and the details rather than just the risk assessment. Perhaps an instrument could be shown that has flaws related to risk assessment and the risk assessor could be asked questions that require them to identify the flaws?