Meta-evaluation of the Comprehensive Needs Assessment to Inform the State of Hawai‘i’s Application for Perkins V Funding

December, 2019

Conducted and prepared by the Curriculum Research & Development Group, at the College of Education, University of Hawai‘i at Mānoa

1 George M. Harrison
Introduction

With Perkins V comes the requirement to conduct a comprehensive local needs assessment and update it at least every two years. This document describes the reflections on the process used in the first year for the purpose of informing the subsequent needs assessments. As a meta-evaluation report, this describes what worked well and what was difficult to achieve and, insofar as these reflections are applicable to subsequent iterations, it serves as a preliminary guide for the evaluations. The intended audience of this report includes evaluators and other personnel who are already familiar with the expectations of the Perkins V comprehensive needs assessment and of the context in Hawai‘i. In addition to drawing from this document, which is specific to our state, evaluators should attend to national-level guides (e.g., ACTE, 2019; Advance CTE, 2019b) and to other documents such as those provided by other states’ needs-assessment guides and reflections as they become available.

The evaluators considered a needs assessment as a systematic collection and interpretation of data to inform decisions or judgments about gaps between the present and desired condition. Included in interpretations are judgments about what is functioning well and what some of the root causes may be for gaps that either might already have been filled or that need attention. In the 2019 project, we were unable to accomplish a clear explanation of the present and desired conditions in all of the categories defined by Perkins V. Much of the needs assessment report attempted to identify what was working well and what needed attention but there were challenges in identifying the desired conditions as well as in identifying the criteria for analyzing what the present and desired conditions are. For example, with regard to program quality, it was difficult to identify a level of quality on specific criteria that would constitute the desired degree of quality. It was also clear that definitions of quality had not yet been agreed upon or understood by the participants. The root-cause analysis was not developed to the extent that we could draw conclusions about causes for gaps. The next needs assessment has an opportunity to expand the timeline and conduct at least two waves of data, with the second wave allocated to these deeper questions.

This bulk of this document addresses the data collection procedures. It also includes ideas about how the next needs assessment can approach stakeholder engagement (whom to talk with), what to assess, and a timeline for data collection.

Data Collection Procedures

Focus Group Interviews

What worked well. In general, it was easy to obtain the buy-in of high-school teachers and university deans to participate in the needs assessment process. We observed that for the most part, these participants sought to be heard. For some participants, there were frustrations that they sought to express; for others, there were experienced perspectives that they sought to share. Because we were an outside entity, with no direct stakes in any particular programs at either the secondary or post-secondary level, we were received—or so it seemed—with openness and, perhaps, a level of candor.

2 In Advance CTE’s (2019) document, these types of information include (a) student performance data on the core performance indicators; (b) programs’ size, scope, and quality; (c) programs’ labor market alignment; (d) progress towards implementing CTE programs and programs of study; (e) recruitment, retention, and training of faculty and staff; and (f) progress towards improving access and equity. In the guide to the law (Brustein, Castillo, Fridman, Passafaro, & Spillan, n.d.), these are listed as five types of information, with the size, scope, quality, and alignment constituting information about CTE programs offered by the local eligible recipients (pp. 39–40).
that might not have been present if we were needs-assessment personnel from their specific schools or colleges.

Contacting the CTE district resource teachers through their scheduled DOE meeting with Troy Sueoka worked well. From this first encounter, the needs assessment team was able to follow-up with individual resource teachers and schedule meeting times with them and the teachers in their districts. This sizes of the focus groups in districts on O‘ahu and Maui were manageable. Contacting the post-secondary deans through Nic Estrella was also easy; we arranged a focus-group session to be held during one of their monthly meetings. The size was slightly too large, but every person participated.

The information from non-CTE teachers was valuable for revealing attitudes and understandings in the system in which CTE teachers operate. The focus groups also included students who were candid about how they became interested in the careers they were seeking and their interactions with the school staff. The focus groups with the non-profit personnel also revealed some perspectives about school-workforce alignment.

The in-person focus-group sessions with the CTE teachers in each of the districts resulted in rich data with back-and-forth discussions and contributions from all or nearly all of the participants. The audio recordings in these sessions tended to be of good quality. Two audio recorders were used on different sides of the discussion table. On at least one occasion, one of the audio recorders began to fail because of electrical interference; having a second recorder preserved the discussion.

The transcription service, Rev.com, had a quick turnaround, though with some files, the transcribers had difficulty with Hawaiian words and Hawaiian Creole English (i.e., Pidgin). This required extra time to review the transcriptions and repair the mis-transcribed portions of the speech before subjecting the transcripts to analysis in NVivo. Without the initial transcriptions, however, this process would have taken much more time.

**Expectations about stakeholder contributions.** For the most part, the teachers, counselors, and principals were eager to contribute. As was expected, each stakeholder group’s contributions reflected their concerns.

With the principals, for example, the topics of their focus-group comments gravitated toward structural and funding concerns: In the academy principals focus group, structures and funding were the most frequently coded themes (21% and 13% of the codes respectively). One principal was very outspoken and expressed a desire for more flexibility in how funding is allocated, which makes sense given their role in securing as many resources to their school as possible and their perceptions of the limitations in some classifications of costs. Another principal expressed a strong desire for structural change—for the state to provide a mechanism for principals to work together to enact changes while also permitting greater flexibility in how they manage state resources. They expressed an expectation of similar structures to those that were introduced to them via the academy model, citing the model they learned about from Jay Steele. Some principals sought for us—the evaluators—to explain to them why the state’s structures were in their present form and why the state was not moving forward in this work. We had to clarify that our role was to learn about the current conditions and we quickly realized our limitations in being prepared to engage in deep discussion about these conditions. It was also difficult for us to extract from the principals what informed their decisions about which programs to allocate more support to in their schools. The most common answer was that the students’ enrollment determined program sizes. We still do not know the extent to which, if any, principals’ use workforce needs or equity-based concerns in their decisions about how to allocate resources among their programs.
Teachers and counselors were more amenable to sharing their experiences, their perceptions of the students’ experiences, and what was going on in their programs. Compared to the principal focus group, the teacher focus groups were much easier to facilitate because the teachers followed the facilitators’ prompts whereas the principals’ contributions reflected an agenda to effect change in the state policies.

**What did not work well.** The online focus-group session with teachers from districts on Kaua‘i and Hawai‘i did not work well. For starters, it was difficult to hear what some participants were saying because of their audio settings or connectivity issues. In some cases, there was excessive background noise as well. Although a resource teacher did her best to recast questions and responses, these conditions prevented the session from having the defining characteristics of a focus group. The back-and-forth interactions were suppressed by these technical difficulties and much of the conversation involved requests for clarification rather than in-depth exploration of varying viewpoints on the topics that were introduced.

Secondly, there were too many participants in the sessions. This was also the case for the sessions with the academy principals, the counselors, and the students. The large size weakened the opportunity for everyone to share their thoughts. With some sessions, particularly the online Kaua‘i–Hawai‘i session, we sensed that there was hesitancy for back-and-forth discussions once a topic was introduced. In these sessions, there were several people who did not contribute to the conversation. It was simply impossible for the facilitators to elicit voices from all members. The audio recordings in these large sessions were of lower quality than the smaller sessions that had been conducted around a table because some people’s voices were distant from the microphone and not easily heard, resulting in transcription problems and error in the interpretation of the participants’ contributions.

Another challenge was scheduling. With one of the districts, the CTE resource teacher fell ill and was unable to respond to requests before the teachers conducted their final meeting of the academic year. The academic year was coming to a close and the resource teachers would not meet again until months later. As a result, we administered surveys to these participants, with the same prompts as those in the focus-group protocols. The interactive characteristic of focus groups was not, therefore, part of the process.

The most salient root cause for these did-not-work-well conditions was scheduling. Because the needs assessment work began in the spring semester with a small team, each with 20 hours a week or less allocated to first learning about Perkins V requirements and then to identifying what the needs assessment should capture, the focus-group data collection was given inadequate preparation time. As a result, the focus-group meetings were arranged toward the end of the academic year, when many other meetings were occurring.

During all of the focus-group sessions, we were unable to cover all of the Perkins V categories of information that are required by the law (Advance CTE, 2019a). We found that the teachers tended to be most vocal about the conditions related to recruitment, retention, and training of faculty and staff as well as about program size, which tended to relate to staffing and resources. The prompts about the use of student performance data tended to lead to the least amount of discussion and we typically did not have adequate time to address progress towards implementing CTE programs. Because the focus-group sessions are limited by time and participants’ attention spans, the next needs assessment should identify which of these types of information are most important for capturing in focus groups and which to allocate to other data collection methods. For those types of information not amenable to focus groups, survey or document analysis methods may be more appropriate. Furthermore, focus-groups for administrators, counselors, and teachers may differ in which types of information are given more floor time, though some overlap may be needed for triangulation across stakeholders. Alternatively, follow-up focus groups can be conducted after an initial wave of data collection, though that will require more
transcription and coding. Multi-wave interviewing, however, will be valuable for root-cause analysis because follow-up sessions can ask not only for verification but for any additional evidence that participants may have observed since the first interview.

Representation of the stakeholder populations was a challenge, particularly with students, teachers who did not identify as CTE teachers, and principals. In part, this was because we gained access to the students through a non-DOE sponsored event to avoid delays in DOE approval for contacting students. The students at that event were very likely highly motivated, as they were voluntarily attending a STEM conference. Those students did not represent student special populations who were in need of equity-based attention. The non-CTE teachers were selected through a local non-profit organization, which was valuable for circumventing the DOE approval process; however, they were teachers who were attending another meeting having to do with becoming active in policy and therefore not necessarily a representative sample of the non-CTE teachers in the state. The principals we met with were also not a representative sample of the state’s principals because they were at an academy principal meeting. The views held by non-academy principals likely differed from those held by academy principals. Subsequent needs assessments can learn from our limitation and achieve better representation of students, principals, and non-CTE teachers.

Other stakeholders should be interviewed in focus groups or with some other method. The 2019 needs assessment did not conduct focus groups with representatives of many of the populations that are deemed important for Perkins V, such as local or state workforce development board representatives, the post-secondary faculty population, parents of HIDOE students, the native Hawaiian community, and special population representatives. The next needs assessment can allocate resources to establishing connections with representatives of these groups and scheduling focus-group sessions or surveys.

**What can be done next.** In subsequent needs-assessment focus groups, the evaluators can do the following to achieve better results:

a. The focus-group meetings should be arranged months ahead of time and be conducted during times of the year when there are fewer responsibilities for participants and with enough time in advance that canceled interview sessions can be rescheduled.

b. The focus-groups should be restricted in size, perhaps no larger than seven participants at a time.
   i. For example, arrange focus groups with smaller groups of counselors, principals, and students.

c. Continue to use at least two audio-recording devices.

d. Prioritize the content in the focus groups. This topic prioritization might depend on the participant type. It is not possible to engage in meaningful discussion about all six categories of Perkins V information in a single focus-group session.
   i. Principals will likely easily contribute to topics about structures, funding, and size concerns. We probably should have asked more about their decisions for program resource allocation, particularly whether and how they included workforce alignment and equity in those decisions.
   ii. Teachers will likely seek to speak more about size, scope, and recruitment, retention, and training concerns.
   iii. Counselors will likely seek to speak more about scope and work-based learning. We probably should have asked for more about workforce alignment and program quality.

e. Conduct focus-groups with better representation of the populations.
i. For example, include principals from non-academy schools and students from a range of academic motivation.

ii. Include participants that represent stakeholder populations we did not include, such as workforce development representatives, UHCC faculty and counselors, parents of HIDOE students, representatives of the native Hawaiian community, and representatives of special populations.

Document Analysis of Existing Reports, Minutes, and Online Sources

**What worked well.** Among the more informative existing reports were those from the New Skills for Youth Initiative (Council of Chief State School Officers, Advance CTE, & Education Strategy Group, 2016; New Skills for Youth Initiative, 2016) and the Castle Foundation (Castle Foundation, n.d.). The findings in our needs assessment were similar in many ways to those in those reports, so these sources provided some backing to the conclusions in our report. The consolidated annual report data were valuable for assessing the secondary and post-secondary systems’ student performance on core indicators and for the analysis of the non-traditional special population. Some data will likely continue to be elusive, such as post-graduation employment data in non-traceable settings.

Among the meeting minutes, the CTECAC minutes were the most prevalent in our analyses because they were easily accessible and well organized. Minutes were informative about what topics and problems were salient and about some of the efforts and opportunities in the system but it was difficult to tell if the activities were subsequently carried out and what their effects were; as a result, these minutes did not reveal root causes for some of the problems. Additionally, the minutes provided little depth to aid in estimating the system’s current status on the Perkins V categories of information. The meeting notes from the 12 meetings—including with Hawai‘i P–20, Jobs For the Future, CTE working group meetings, and communications meetings—were informative for revealing general concerns and for identify key participants in the state’s system. These were very informative for learning about the system and about plausible system-level root causes, but not valuable for evaluating program quality and equity.

The most informative online documentation was in the colleges’ Annual Report of Program Data (ARPD) system. These provided evidence about how the local programs in each college are collecting data and reporting on their programs’ quality and alignment. The ARPD system was valuable for examining individual programs.

There was also DOE documentation about some of the standards for the programs of study as well as a CTE resource teacher handbook. These revealed that some pathways were more developed than others. For addressing scope, this type of documentation was valuable, though we lacked the resources to examine every program of study and we did not have a specific set of criteria for rating what each program’s status was or what the ideal scope would be.

For the analysis, NVivo software worked well because we were able to code different sources, including web pages, images, and PDFs. With PDFs and other text documents, specific words were included in the codes, whereas with a website, the part of the page that included the words did not appear in the coding. This was annoying when we interpreted clusters of codes.

**What did not work well.** One expectation we had based on the language of Perkins V guiding documents was that we would be able to identify data from existing documents and other records and then estimate, for some of the categories of Perkins V information, what our current status is and where
we would like to be (thereby providing an assessment of needs).\(^3\) We failed to find a way to achieve this with satisfaction. One of the expectations in the Perkins V law is for large endeavors (WIOA, ESSA, and so forth) to be aligned (Brustein, Castillo, Fridman, Passafaro, & Spillan, n.d.). Related to this, it is expected that data be made available across endeavors for use in achieving desired results. With this, the implication is that the needs assessment could use the reports of aggregated data in these existing documents—whether they be from Perkins or other endeavors—for inferring the current and desired status in ways that can inform individual programs or sites. We were unable to achieve this at the program level, though it may have been possible had we known what variables were available in the source data for those reports. One thing that would have helped is a code-book for such data frames, as the evaluation team could then have read through the descriptors and determined if those data would serve the needs assessment. However, precluding this was the need to identify what desired content criteria we should be looking for in these documents.

Another challenge we experienced was in managing the plethora of plausible reports and documents that might contribute to the document analysis. Some documents were clearly valuable, including the documents and reports by the DOE and the UHCC system. Those of the Hawai‘i Workforce Development Council were informative for very general understanding of the state system. There were some documents that revealed workforce trends in different locales in the state that might have proven useful for evaluating alignment of local programs. Those of P–20 revealed important information, for instance about work-based learning opportunities and about graduation rates of CTE students. We found ourselves being concerned about the sources of some nationally distributed documents. For example, a report from Credentials Matter seemed to be funded by partisan think tanks. As evaluators, we needed to be cognizant of the implied intended purposes of the documents. For some reports, the purpose may be to support student success in career and technical education that leads graduates to attain their desired lifestyles and secure high-wage, high-skill, and in-demand employment throughout their adult lives. For other reports, the purpose may be to support economic growth, some of which may be for economic health of the local economy. For still others, the interest may be for strengthening business. Though these purposes all comport with Perkins V, the weight we allocated to the data in those documents was affected by our perceptions of how closely the interests addressed the needs of the students.

What can be done next. If our evaluation team were to do this needs assessment again, I would ask that we classify which types of data are pertinent to document analysis. Data about scope and student performance seemed to be captured in the documentation. In the ARPD, data about workforce alignment were present. It also might be collect data at two levels: at the general eligible-recipient level for identifying needs and assets and then at the local level for identifying root causes.

Subsequent needs assessments can ask for a list of variables, like a survey code-book, from the eligible recipients contributing to the consolidated annual reports. If such a list is possible, and if the data can be dis-aggregated by program, some of the needs for each program might be identifiable. This data identification task might be a state-level needs assessment role while the individual programs are analyzed using these data, in a manner similar to how individual post-secondary programs evaluate their progress in the ARPD system. Prior to that work, however, there is a need for clarity in what constitutes the desired level of size, scope, quality, and alignment. Future needs assessment work might also seek to identify whether other documents and minutes exist that can reveal patterns in the system and provide indicators of what the current status is in comparison to the desired status.

---

\(^3\) As listed above, these are (a) student performance data; (b) size, scope & quality; (c) labor market alignment; (d) progress towards implementing CTE programs and programs of study; (e) recruitment, retention and training of faculty and staff; and (f) progress towards improving access and equity.
Surveys

**What worked well.** There were few benefits observed in using surveys in this project. Perhaps one exception was the capacity for the online surveys to be perceived by participants as anonymous, as we did collect some seemingly candid data. This can be because there is lower social-desirability bias and greater willingness to share outspoken views in response to anonymous constructed-response prompts. However, this can also result in low response rates. In one of our survey administrations, of the approximately 6 participants invited by the district resource teacher, only 3 responses were recorded. Another obvious benefit to surveys is the capacity to quantify the results and compare the current status — on the questionnaire response scales — to a desired cut score on that same scale. Such information provides results about the needs that are easy to interpret and communicate.

**What did not work well.** The 2019 needs assessment used surveys for two purposes: (a) as a replacement for a focus-group that could not be scheduled and (b) to request CTE teachers’ perceptions of the value of the NAPE information after a training session on NAPE dashboards. These were very small data-collection endeavors, with low stakes, and not even good for that use because the response rate was so low. Low response rates are not a problem if we can be sure that those who chose to participate are no different on the variables we seek to learn about than the participants who chose not to participate. In this sense, the primary threat to the validity of the interpretations of the survey data was non-response error. Unfortunately, there is no easy way of detecting this type of error (which is why evaluators tend to focus on a response rate index, which when very high means that non-response bias is low; however, the inverse is not necessarily true).

Initially, the lead evaluator had intended to use surveys to measure how well programs met desired criteria in size, scope, and quality. However, a problem emerged in determining what content to include in surveys. As the lead evaluator began developing a survey to measure perceptions of programs’ scopes, he found that the number of variables sought in the needs assessment far exceeded the capacity for a single survey to provide quality data to address every criterion. It was soon discovered that for the survey to be successful, the most important criteria for evaluating programs need to be identified. Otherwise, the instrument would become too long and risk high drop-off rates and high degrees of participant satisficing because of fatigue. Worse, teachers who will have suffered through the instrument may complain to their colleagues about the experience, resulting in subsequent teachers’ avoidance of the survey and low response rates, thereby potentially increasing non-response error.

Additionally, some survey questions may not be appropriate for a survey unless anonymity can be ensured because acquiring candid responses may be challenging when hard questions are asked, such as whether a program should be eliminated. If a single teacher is responding about a single program, anonymity is clearly absent. The stakes are high for teachers who will lose their job if their program gets cut. There is also a concern among some DOE staff of being held accountable for making statements that might not represent the DOE’s views, as one participant candidate—who opted not to participate—stated.

**What can be done next.** Surveys should be kept short. Roughly, a 10-minute survey may be appropriate, whereas a 15-minute one would not, unless the questions are interesting to the participants and their attention can be sustained. This requires a prioritization of the content that will go into a survey. The people who are willing to allocate time and attention to a survey may be very different from those who are less engaged. Ideally, participation will be highly encouraged, perhaps by asking participants to

---

4 In survey research, satisficing refers to participants’ engaging with survey questions with low levels of attention, which results in higher measurement error and, therefore, weaker validity in the claims about the phenomenon being measured. In contrast, optimizing refers to participants’ careful engagement to provide a response that optimally represents their true status on the concept being asked about.
complete it during a required meeting (and not right before or after a break). Evaluators may ask which criteria the state and the locals can live without collecting, as well as which variables are more appropriate using other methods. If there is a need to triangulate data across instruments or participants, the survey invitations should not be anonymous and can be customized and written directly to the individuals to encourage participants to respond. For topics that are prone to social desirability, a separate anonymous survey may serve as a supplement, though the response rate may be lower (however, imposing it as an activity in a required meeting may curtail that validity threat). Surveys can provide quantitative data and reveal the gap between perceived current conditions and the ideal state. They can only measure so much, though.

**Summary of Data Collection Procedures**

With regard to data collection procedures, several decisions are required. For instance, we need to determine which content to measure and what mode of data collection is most appropriate for that content. For decisions in prioritizing and operationalizing the criteria used to assess the needs in program size, scope, quality, and alignment, a flow chart may prove useful. The flowchart below serves as this initial attempt, though future evaluators may find cause to revise it or take an entirely different approach to decisions about what to measure and how to measure them. In this most recent year, the criteria for size, scope, and quality were being revised. One additional step may be to rank these components, or criteria, to determine what the evaluation can live without measuring and what must be measured at all costs. This triage process may be tricky because we also seek to examine root causes for low scores on some criteria. For instance, a particular program may have few or no opportunities for its students to acquire work-based learning experiences; is this because the program is so cutting edge that such experiences are unavailable or is it because of the rural location of the program’s students? A survey may not be appropriate for examining root causes. At the same time, focus-groups are limited in the amount of time and the number of topics that can be discussed.

Additionally, it is not enough to have a checklist of statements constitute the feature (such as quality); there is a need for a range to be implied in that criterion. For instance, if the criterion is to “engage employers,” there needs to be a way to track the growth of the program in its employer engagement; in this case, there is an implied metric of frequency (or perhaps quality) in how employers are engaged. Engagement might not be the correct criterion, either, because some programs and sites experience more difficulty engaging local workforce entities than other programs and sites. The implied metric, therefore, seems unfair when we use it to locate a desired status to which we can compare a current status because programs and sites differ. Nonetheless, the same metric should be applied across sites for us to understand the root causes and to direct resources, not for penalizing poor performance.

---

5 In addressing survey length, what we’re really after is how to optimize the instrument to get the most accurate responses from the participants. If the respondents are very engaged with the survey because it is interesting (and this interest level is not differential across levels in the construct we’re intending to measure), we can have a survey that takes a long time (or has many questions) but making it interesting is not easy. In other words, there is no magic number of questions or specified length of time. A boring survey should be short. An interesting one can be long.
Decisions in Operationalizing Program Size, Scope, Quality, and Alignment

- Are the criteria appropriate for the state and local priorities? Yes → Request or develop appropriate criteria. No → Revise into working criteria that adequately represent the official criteria, or share possibilities with the state for revising the official criteria.

- For each criterion, will there likely be strong bias due to survey self-report? Yes → Flag the criterion for use in surveys. No → Flag the criterion for use in interviews.

- Are the criteria worded to imply a scale from low to high (or a present or absent condition)? Yes → For criteria that require decision rules for judging how good is good enough, are there obvious rules? No → Facilitate a standard-setting session to establish cut scores on the scale from the instrument(s).

- Are these criteria ranked in priority? Yes → Have the criteria ranked based on state and local input. No → For each criterion, can existing data be used to evaluate a location on a rating scale? Yes → Flag the criterion for use in a rating scale.

- For each criterion, can existing data be used to evaluate a location on a rating scale? Yes → Review instruments for validity based on the content and for feasibility in using the data for making decisions about programs' status on the feature. Return to appropriate step in flow chart, or commit to data collection.

- Identify the best method for each criterion. If no method is appropriate, investigate alternative data collection methods.

- Identify criteria that require triangulation, and identify the secondary and tertiary methods.

Draft instruments (rating scale protocols, survey instruments, and interview protocols), with the length of each limited to feasible administration, participant optimizing, and attention to the criteria ranking.
**Whom to Talk With: Engaging Stakeholders**

In the 2019 needs assessment, we faced the challenge of learning about the system, planning the needs assessment, identifying and contacting stakeholders, and conducting the needs assessment in a short amount of time and with a staff of one to three people, depending on the month, all of whom were part-time on this project. Engaging representatives of various types of stakeholders is listed as a priority in the Perkins V guides (ACTE, 2019; Advance CTE, 2019a). For the 2019 needs assessment at the secondary level, we placed CTE teachers at the top of this stakeholder list, with principals and counselors next, and students after that; at the post-secondary level, the deans were most accessible. Whereas high school CTE teachers were easily identified through CTE district resources teachers (who were contacted through the DOE’s CTE office), nearly all instructional faculty at the post-secondary level conduct work in programs that are classifiable as Perkins funded, requiring a different method of teacher selection for focus groups. We did not interview post-secondary faculty because we were unable to identify which programs might merit more needs-assessment attention than others. The next needs assessment may choose to acquire a sampling frame of all UHCC programs that receive Perkins funds, perhaps stratified by pathway, and randomly sample a number of programs from those strata, subsequently contacting one faculty member from each sampled program to interview or survey. As stated above in our section on focus group interviews, in addition to the stakeholders included in the present needs assessment, stakeholders that should be included in the next round include the following:

- a. parents and students who represent a range of backgrounds (the 2019 needs assessment included students who attended a STEM conference, which was not representative of the state’s students);
- b. representatives of special populations (the 2019 needs assessment included non-CTE teachers, several of whom worked in special education, but no representatives of the other special population groups were successfully included, though attempts were made);
- c. representatives of the native Hawaiian community;
- d. representatives of the state or local workforce development boards (the 2019 needs assessment included representatives of non-profit organizations, not of the workforce); and
- e. post-secondary counselors and program instructors (the 2019 assessment included only the deans—the perspectives of these other participants would likely reveal more about the present system).

**Getting Locals to Collect and Analyze Needs-Assessment Data**

A noticeable weakness of this first needs assessment was its absence of the “local”. Local sites have greater familiarity with their programs and greater access to sources of evidence, including documents of the curricula, student demographics and conditions, and assessment results, as well as teachers, students, business partners, and community members. If it is possible for the instruments to be developed enough time in advance, the evaluation team can ask for local needs assessors to be identified within each DOE geographic district and each college. A straightforward rating scale rubric for evaluating program quality, scope, and workforce alignment may prove useful. Short surveys may also be worthwhile. The ARPD system at the post-secondary level already provides one means of accomplishing this. Still, at both the post-secondary and secondary level, there may be a need for instruments and processes. Other states may have models we can refer to for this process.
Ideally, the evaluators will have a template for the locals to conduct needs assessments of their programs and programs of study. This should be guided by the prioritized types of data. There are also guidelines from ACTE (2019), stating that locals should examine programs based on how well they

- incorporate the state’s ESSA standards (secondary only);
- address academic skills, technical skills and knowledge, and professional skills;
- address alignment of industry needs;
- include progressions of industry or pathway education from broad to specific;
- permit multiple entry and exit points for students into industry, such as through credentials at multiple levels; and
- include opportunity for students to eventually achieve a post-secondary credential.

**What to Assess**

In the 2019 needs assessment, although we did not begin with strong assumptions about the local secondary schools’ definitions of criteria for quality, size, scope, and alignment, we found that there were discrepancies between what the locals considered to be salient and important in these criteria and what appears to be emphasized in Perkins V. Additionally, the size, scope, and quality criteria that were developed for Perkins IV (provided in Appendix D of the needs assessment report) included a list of statements that were difficult to operationalize into an instrument that measured that entire set of content. For instance, the scope and quality criteria each included at least 13 statements, many of which were multidimensional. After our initial attempts to develop a survey to measure the scope and quality criteria, we abandoned the process because the instrument became too lengthy and would very likely invite satisficing or worse, survey dropouts and non-response bias, which are already problems with teachers and professionals who have been extensively surveyed. Another problem with operationalizing content criteria were that they seemed to overlap; for example, the scope criteria statements largely addressed quality and alignment. It is understandable that scope includes alignment because curricula and work-based learning activities—both of which are part of scope—should be informed by workforce alignment. However, what exactly does our state seek to know about programs’ desired and observed status in workforce alignment and how is that distinct from what we seek to learn about programs’ status in scope criteria?

Additionally, the quality criteria were mostly procedural checklist items, without an underlying metric for judging distance between observed and desired status. If we seek to track programs’ growth or assess the discrepancy between current and desired status, a scale is desired. Checklists can provide this information aggregated across programs if our desired status is a percentage of programs meeting that particular criterion. It is also possible to treat checklist criteria as dichotomously scored items (i.e., 0 points for not meeting and 1 point for meeting the condition) and establish each individual program’s desired “score” as a number of content criteria out of the total that is achieved (similar to a classroom spelling test). However, that procedure assumes that every item has an equal value when we sum up those met conditions out of the total. It may be necessary to rank the importance or value of the content criteria within each type of information for two reasons: (a) to provide an appropriate measurable estimate of the current status of each program, and (b) to identify what information we can afford not to measure, given the limitations in what can be included in the needs-assessment instruments.

Fortunately, with Perkins V, the definitions for size, scope, and quality are being revisited. One way in which the next needs assessment can be made more systematic in assessing not only these three
criteria but also those within the categories of information in Perkins V is to break the process down into several steps:

1. **identify prioritized content criteria, within each type of information, as agreed upon by both the state and locals;**
2. **establish how good is good enough (benchmark criteria) on the most important criteria;**
3. **conduct a first wave of data collection, targeting the present condition (for subsequent comparison with the benchmark criteria to estimate the gaps);**
4. **conduct a preliminary analysis on that data to identify gaps, questions to drive the root-cause analysis, and any other issues; and**
5. **conduct a second wave of data collection for targeting root causes, addressing unanticipated issues in the first data collection, and triangulating findings.**

A needs assessment data-collection timeline, such as the one displayed at the end of this report, can direct this process.

**Identifying prioritized criteria.** The Advance CTE guide for the comprehensive local needs assessment (2019b) suggests we consider what our state’s priorities are for the needs assessment. Although that guide suggests these are in addition to the requirements of the law, within each of the required categories of the law (listed in an earlier footnote), there are still criteria that our state defines (such as size, scope, quality, and alignment). Entailed in this process is a prioritization of these content criteria (within each of the five or six types of information required by the law). The next needs assessment may ask for the criteria within each of the four categories of quality, scope, size, and alignment to be ranked in order of importance prior to preparing the data collection. An alternative to ranking each individual criterion may be to classify them in ordered buckets of importance.

Regardless of the ranking approach, these criteria should be understood by both the local and state personnel. A foreseeable continuing challenge will be that the priorities of the locals will differ from those of the state (Advance CTE, 2019b). The next iteration of the needs assessment will do well to include representatives of the local stakeholders in identifying a subset of most-important criteria within each of the four categories and perhaps within the other four types of information. Because the definitions for size, scope, and quality—and presumably for alignment—are being rewritten in the current year, there will be a revised list of criteria to frame the discussion among representatives from the state and local stakeholders as well as from the needs assessment team. One possible approach is to contact these representatives several months in advance of the needs assessment data collection and ask them to rank these criteria, perhaps with an accompanying discussion of the consequences of each criterion to help to determine importance. Subsequently, the evaluation team can identify which criteria merit the most attention in the needs-assessment instruments for data collection. With content criteria that rank highest, triangulation will strengthen the validity of the claims the needs assessment makes from the data.

**Establishing how good is good enough.** Following this, the evaluation team can identify a scale for each content criterion, such as is often found in education rubrics or survey response scales and within each scale establish the cut-off criterion (or benchmark) for adequacy. This cut-off approach is similar to that used in educational standard setting and must include, again, representatives of the stakeholder groups. As such, the second phase of the needs assessment can include standard-setting sessions. For example, a group of stakeholder representatives can be asked to review each survey question and each document

---

6 As listed above, these other four types of information are student performance data; progress towards implementing CTE programs and programs of study; recruitment, retention and training of faculty and staff; and progress towards improving access and equity.
rating prompt and its response scale (or rating scale) and make a judgment about what the state’s (or local’s) desired and realistic status should be. This provides a benchmark criterion against which to compare the subsequent response data.

Conducting a first wave of data collection. The form of the instruments will depend on the content being assessed. If the content criteria require document analysis, rating scales can be used, particularly for collecting quantitative data. If it requires participants’ reports and should be quantifiable, surveys can be used, though they are prone to many biases, such as social desirability and non-response bias. If in-depth analysis is needed, such as for examining root causes through follow-up probes, interviews or focus-group interviews may be more appropriate as well as subsequent qualitative document analysis to triangulate the findings across types of data sources.

Conducting a preliminary analysis. Following the first wave of data collection, an adequate assessment of the gap between the present condition—infected by the data that are collected—and the desired condition—infected by the benchmark criteria—will likely be achievable. In interpreting these data, questions may arise about root causes or about the validity of the interpretations. Follow-up surveys, document analyses, or focus groups may be needed to investigate more deeply the most important criteria or to verify earlier findings.

Conducting a second wave of data collection. A subsequent step in the data analysis can then attend to root causes as well as to addressing unanticipated concerns about data quality in the first wave. Focus groups and surveys might be conducted with the same participants as in the first wave for member checking or for deeper probing. Alternatively, the evaluators may decide that new participants from the same population may serve these needs better.

Data Collection Timeline

The planning for the needs assessment was carried out from late January through June of the academic year, with the bulk of the data collection carried out between April and June. This was not enough time to fully meet the Perkins V requirements. The timeline draft presented here can be revised for the next needs assessment. This portion represents the data collection. It is based on the assumption that prior work has been done in the state and with the local schools and colleges to establish the content criteria for size, scope, quality, and workforce alignment. Also not displayed is the subsequent data interpretation, the report write-up, and the dissemination of the findings.

Final Comments

With Perkins V comes the requirement to conduct a comprehensive local needs assessment and update it at least every two years. The 2019 needs assessment was a first attempt to examine Perkins V implementation at the system level. The present document provides a formative meta-evaluation of the state’s needs assessment work. It addressed what worked well and what did not work well with the data collection, as well as some suggestions for (a) engaging stakeholders, (b) prioritizing what to assess, and (c) collecting data to achieve a deeper understanding of the causes and to strengthen the validity of the needs-assessment claims.
## Needs Assessment Data Collection Timeline

<table>
<thead>
<tr>
<th>Whom to Talk With</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Identify stakeholder contacts</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contact representatives of stakeholders for focus groups, surveys, and for identifying desired conditions on instruments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identify candidate personnel who can serve as local needs assessors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provide local needs assessors with shortened versions of the instruments and request they collect data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local needs assessors share their data and preliminary interpretations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>What to Assess</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Identify prioritized criteria</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distinguish between the state's and locals' judgments of what should be prioritized</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Identify desired conditions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identify desired conditions, such as in standard-setting sessions that identify cut scores on instruments' ordered scales</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analyze data to this point and identify what to assess in follow-up data collection, such as for root-cause analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Develop follow-up data-collection instruments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conduct final phase of data collection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Document Analysis</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Identify types of content most pertinent to document analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Draft the rating scale criteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Establish desired conditions, such as desired locations on rating scales</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analyze the document data to inform follow-up data collection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identify follow-up rating scale criteria to identify specific themes for triangulation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conduct secondary analysis of specific themes on selected documents as needed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Surveys</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Identify the content of the surveys</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Draft the survey questions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Establish desired conditions on the survey scales</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administer the surveys</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analyze the survey data to inform follow-up data collection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Develop follow-up survey questions, as needed, from earlier analyses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administer follow-up anonymous surveys to selected stakeholders as needed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Focus Group Interviews</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Identify content of focus groups</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Draft focus-group prompts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conduct early phase focus groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analyze the focus-group data to inform follow-up data collection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Update focus-group prompts, such as for root-cause analysis, using data from early phase focus groups and from surveys and document analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conduct later phase focus groups with selected stakeholders as needed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


Council of Chief State School Officers, Advance CTE, & Education Strategy Group. (2016). Hawaii New Skills for Youth Initiative Phase II (pp. 1–44) [Comprehensive Assessment].

New Skills for Youth Initiative. (2016). Lessons learned, challenges faced in advancing career readiness in Hawaii’s schools: A diagnostic assessment of Hawaii’s career readiness program based on the work of the Hawaii academies high schools (pp. 1–31) [Source unknown].