August 2019

Publication 18.25

Leadership Development: Summary of Evaluation Findings From 2018–2019

Department of Research and Evaluation
AUSTIN INDEPENDENT SCHOOL DISTRICT

Executive Summary

This report describes findings from the Austin Independent School District's (AISD) Department of Leadership Development 2018–2019 evaluation. The purposes of the evaluation in 2018–2019 were to update the Department of Leadership Development theory of change (ToC) and logic model and to evaluate the implementation of the newly developed assistant principal (AP) and principal hiring processes designed to select a diverse group of equity-focused, effective novice leaders. This evaluation was designed to help the department meet its goal of creating a leadership pipeline that reflects AISD values and fosters a selection and development mindset. Underlying this goal is the core belief that the creation of this leadership pipeline will lead to improvements in the quality of campus leaders in AISD, as well as improvements in staff retention, student achievement, and campus climate and culture. Therefore, the evaluation efforts in 2018–2019 were executed to help reach this goal.

While the Department of Leadership Development engages in many areas of work, such as AP induction and support and novice principal support (see Figure 2 for a full depiction of the department's work), in 2018–2019 the department focused on the implementation of the new AP and principal hiring processes. These new hiring processes were a multiple-hurdle approach to personnel selection consisting of a HireVue interview, an assessment center, and a campus-based interview. Therefore, the evaluation efforts for the 2018–2019 school year mirror this emphasis on new hiring procedures for APs and principals. The main evaluation tasks from 2018–2019 are as follows:

- Created a theory of change and completed a logic model to more clearly delineate the work, goals, and expected impacts or outcomes the department expects as a result of properly implementing the work and reaching their goal.
- Hosted a workshop that focused on summer 2018 AP HireVue data analyses, AP assessment center data analyses, and best practices for conducting assessment centers and interviews.
- Conducted interviews with current APs and principals to gain perceptual data regarding the new hiring processes.
- Administered surveys to AP and principal applicants to gain and analyze feedback on the new hiring processes.
- Analyzed spring 2019 principal HireVue and assessment center data.

Findings from the evaluation provide evidence that individuals who participated in the preparation programs, such as the Assistant Principal Preparation Program (AP3), Principal Preparation Program (P3), or the Texas State program, were more likely to pass the new hiring processes than those who have not participated in a preparation program. Additionally, perceptions of the new hiring process were mostly positive from both applicants and those hiring applicants. Applicants saw the new hiring processes as generally fair and as processes that accurately assessed their readiness to enter into a campus administrator role. Some recommendations, such as working to increase the diversity of the applicants and provide clear messaging about the new hiring processes, were also voiced.

Table of Contents

List of Figures

Introduction

Purpose of Evaluation

The main purposes of the 2018–2019 evaluation of the Department of Leadership Development were to update the Department of Leadership Development ToC and logic model and to evaluate the implementation of the newly developed AP and principal hiring processes designed to select a diverse group of equity-focused, effective novice leaders. This new hiring process included an online video interviewing tool, HireVue, and an assessment center consisting of several tasks intended to assess candidates' abilities to perform job-related duties.

Department Background and Description

Developing a sustainable and scalable pipeline of outstanding urban school leaders is a vital component of the AISD *Strategic Plan: 2015–2020*. Specifically, the Department of Leadership Development is aligned with Core Belief Two, which states that AISD will create an effective, agile, and responsive organization. Additionally, while not directly reflected as a measure in the 2018–2019 district scorecard, the goals of leadership development also align with Constraint Three, which states that the district should not conduct itself in an inequitable or discriminatory manner. By creating rigorous selection processes that aim to achieve diversity, the Department of Leadership Development is working to meet Constraint Three. The AISD Office of Human Capital is committed to creating a rigorous leadership pipeline that focuses on seeking, finding, preparing, and keeping a diverse group of equity-focused, effective novice principals and assistant principals (APs). There were several major focus areas of leadership development in the 2018–2019 school year.

- Leadership Foundation Programs: A partnership with Texas State University and The University of Texas at Austin, offering opportunities for current teacher leaders to earn advanced degrees or get placed in a residency at an AISD school to further prepare them for a leadership role in AISD
- AP3: A program designed for teachers that focuses on providing specific AISD content and preparing cohort members for campus leadership in an urban setting
- P3: A program aimed at preparing selected APs for principalship in AISD by attending sessions with district leadership to increase their knowledge of AISD initiatives and to understand the significance of their role in creating an effective, agile, and responsive organization
- AP and principal hiring processes: New, multiple-hurdle hiring processes developed in 2017–2018 that involve candidates participating in an online, HireVue interview and an assessment center with the goal of being added to the highly qualified list and considered for AISD AP or principal roles
- New AP induction and support program: Supports in place for new APs, such as an induction program at the beginning of the school year to prepare for their leadership on an AISD campus and ongoing support throughout the year in blended formats and learning series

- New principal induction and support program: Supports in place for all principals new to AISD or new to principalships, such as (a) the 2-day institute in the summer to help prepare them to start the year on their campus as a principal and (b) ongoing support through the novice principal coaching program

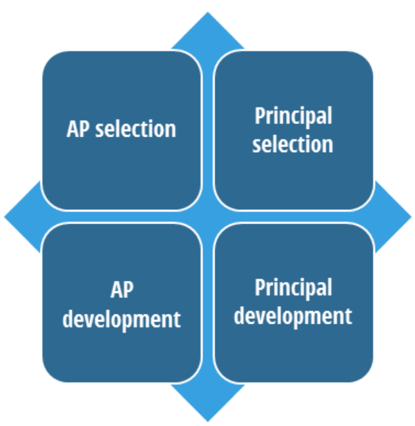
Department Goals

The goal of the Department of Leadership Development is to create a leadership pipeline that reflects AISD values and fosters a selection and development mindset. Underlying this goal is the belief that the creation of this leadership pipeline will lead to improvements in the quality of campus leaders in AISD, as well as to improvements in staff retention, student achievement, and campus climate and culture.

Department Structure

The Department of Leadership Development is overseen by the executive director of Talent Acquisition and Development and led by the director of the Department of Leadership Development, with support from the leadership development team. The work of the department is oriented toward four main areas: AP selection, AP development, principal selection, and principal development (Figure 1).

Figure 1.
Four Main Areas of Leadership Development Work



Description of Evaluation

Overview

The goal of the 2018–2019 leadership development evaluation was to update the department's ToC and logic model, evaluate the new hiring processes, and provide preliminary assessments of candidates who passed through the new processes. Three major questions guided the evaluation work:

1. How was the department conceptualized?
2. Was the hiring process operating as designed or intended?
3. Was the hiring process producing what was intended?

Understanding the conceptualization of work in the department (i.e., question 1) required special attention to understanding the design and intentions of the new hiring process. To comprehensively summarize and describe the process, a ToC and logic model were developed. In this process, information regarding the work of the department was gathered from the Department of Leadership Development website and from meetings with the leadership development team

To address question 2, the evaluation examined operation-oriented evidence and data. Operation is a question of implementation—that is, what evidence indicated that all the pieces and parts were in place for the new hiring process and operating effectively? For the purposes of the 2018–2019 evaluation, much of the focus was on assessing the AP and principal hiring processes, as these were the newer pieces of work for the department. A review of best practices and common heuristics for scoring interview items and assessment center tasks as well as for conducting interviews and assessment centers was provided. Feedback surveys were also provided to applicants who completed the HireVue interview and to applicants who completed the assessment center. Information related to both processes was collected in the feedback surveys and will help to address question 2.

To address question 3, the evaluation focused on understanding if the hiring process was performing as intended. This was simply a question of production, not of impact—that is, what evidence indicated that the department produced what it was supposed to? With the goal of having a hiring process that selects a diverse group of equity-focused, effective novice leaders, the evaluation examined pass-through rates of applicants, with special attention to pass-through rates for certain demographic or other characteristics (e.g., race/ethnicity, gender, program participation). To understand how the new system of hiring performed (e.g., Were items/tasks too easy, too difficult? Were items differentiating participant performance?), data related to item performance on the HireVue interview items as well as task performance in the assessment center were analyzed. Furthermore, qualitative data provided in interviews with AP candidates and hiring principals were analyzed.

Evaluation Methods

As stated, the main purposes of the 2018–2019 evaluation of the Department of Leadership Development were to update the department's ToC and logic model and to evaluate the implementation of the newly implemented AP and principal hiring processes as methods for selecting a diverse group of equity-focused, effective novice leaders. To guide our evaluation, several more detailed questions related to questions 1 through 3 were developed:

1. How was the department conceptualized?
    a. What was the goal of the department? More specifically, what did we want to change as a direct result of the implementation of the work in the department?
    b. In the big picture, what was the point of the department's work? More specifically, if the desired change happened, then what outcome(s) did we hope would improve in the district or on campuses?
    c. What were the major categories of implementation work that contributed to accomplishing the goal?
2. Were the hiring processes operating as designed or intended?
    a. Did hiring processes reflect best practices for item development, interviewing, scoring, and conducting assessment centers?
    b. What were applicants' perceptions of the HireVue and assessment center processes?
3. Were the hiring processes producing what was intended?
    a. What were the pass-through rates of applicants in both HireVue and the assessment center?
    b. What were the pass-through rates for those who participated in a program in comparison to those who did not participate in a program?
    c. What were the perceptions of the AP highly qualified list and the new hiring processes?

Leadership Development Conceptualization Process

After each of the ideas from question 1 were articulated, they were further framed into a programmatic ToC statement. The ToC statement took the form of *If-Then-Therefore*. If we achieve the goal of leadership development, then we will enable the proposed positive outcomes. Therefore, we should engage in the implementation work of the department. The ToC statement was conceptually translated into work, goal, and outcome constructs and organized into a ToC graphic. The overall ToC document was operationalized into sets of measures specific to the implementation, the goal, and the outcomes.

Hiring Process Operations

To determine if the new hiring process was operating as intended, several pieces of evidence were gathered. First, information was gathered from the leadership development team regarding the HireVue items, scoring processes for those items, information related to the assessment center tasks, and scoring processes for the tasks. Based on this information, recommendations could be made regarding the streamlining of scoring processes and structure of items and tasks.

Explanation of ToCs and Logic Models

The purpose of a logic model is to provide stakeholders with a road map describing the sequencing of work and events that are needed in order to expect to achieve the desired outcomes (W.K. Kellogg Foundation, 2004). Mapping out the different areas of work, the goals, and the desired outcomes of that work helps to visually see the alignment, or misalignment, between different pieces of work and the expected outcomes. A logic model tends to be very detailed, inclusive of all the areas of work, how we can measure the works' completion, and how we can determine if we are meeting our goals and achieving our outcomes. As work, goals, and desired outcomes change, so too can the logic model.

A theory of change, or ToC, is another, similar way to articulate the expected outcomes and the activities needed to achieve the desired outcomes. ToCs can sharpen the planning and implementation of an initiative, can highlight necessary areas for measurement or data collection, and help to delineate goals or desired outcomes at the outset of the initiative (Connell & Kubisch, 1998). For the purposes of this report, we created both a ToC statement and a ToC graphic (Figure 2) to display the work, goals, and desired outcomes of the Department of Leadership Development.

Feedback surveys also were administered to both principal and AP candidates regarding their experiences in HireVue and in the assessment center. Thirty principal candidates and 23 AP candidates responded to the assessment center feedback surveys, while 23 principals and 78 APs responded to the HireVue feedback surveys. Although these surveys contained many questions, those most pertinent to the new hiring processes concerned perceptions of fairness, task or item relevance, scoring, connection between tasks, and the use and helpfulness of the toolkit.

Hiring Process Outputs

Pass-through rates were examined for AP candidates who completed their HireVue (n = 350) and assessment center (n = 123) in May through July 2018. From our analyses, we were able to ascertain how candidate groups differed in their progression through the hiring process and how candidates performed throughout each step of the hiring process. For the purposes of this report, pass-through rates comparing those in a preparation program to those not in a preparation program are included. Pass-through rates using more sensitive, demographic information (such as pass-through rates by race/ethnicity or by gender) are not included in this report but have been shared directly with the leadership development team.

Scoring distributions also were analyzed for both HireVue items and assessment center tasks overall and by item/task. In total, five HireVue items and four assessment center tasks were rated on a scale from 1 through 4. The assessment center tasks also had subcomponents that applicants were rated on within each of the four tasks. For the purposes of this report, these score distributions have not been included, but have been shared directly with the leadership development team.

Lastly, feedback was gathered from five hiring principals to ascertain their perceptions of the highly qualified list (containing AP candidates who successfully completed the HireVue interview and assessment center) and of the logistics of the new hiring process. Additionally, four AP interviews provided perceptual data related to the performance of the hiring process.

Results of Evaluation

Conceptualization of Leadership Development Implementation and Goals

Through collaboration with the leadership development team, the work of the department was reconceptualized. A new purpose statement was developed, the implementation work was reorganized and added to, a goal statement was agreed upon, distal district outcomes were identified as possible areas of impact, and a new ToC statement underlying the Department of Leadership Development was developed. A graphical representation of the ToC was also created (Figure 2), as well as a logic model containing more details related to the implementation work, process outputs and outcomes, and performance measures. The logic model is available upon request.

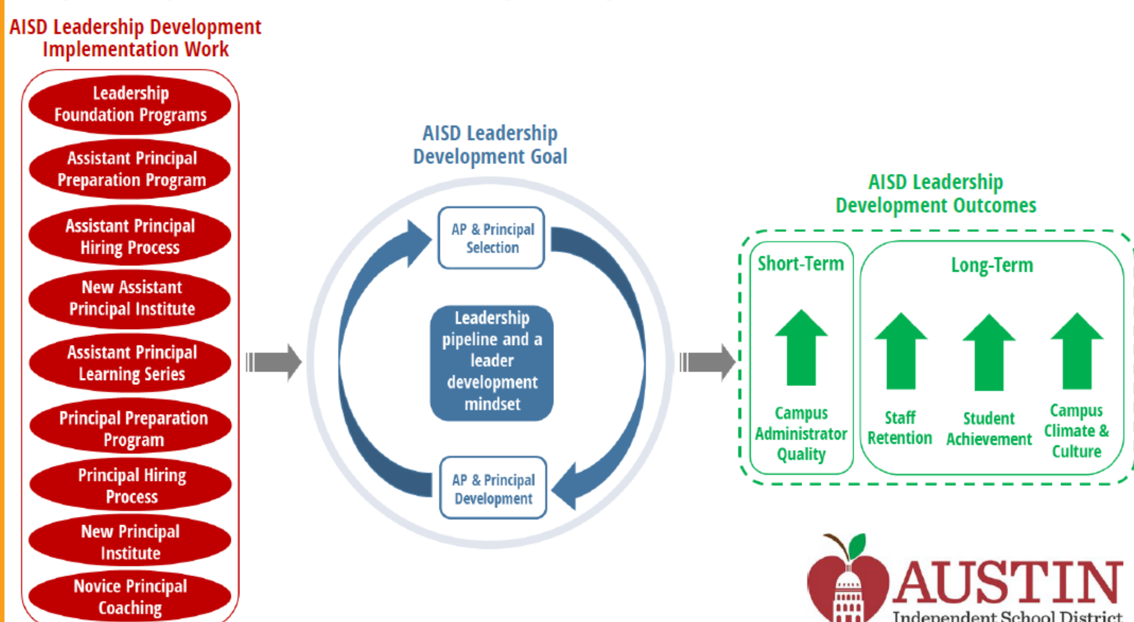
Statement of purpose:

> The purpose of the Department of Leadership Development is to develop a sustainable and scalable pipeline of outstanding urban school leaders. AISD Human Capital Services is committed to creating a rigorous and scalable leadership pipeline that focuses on seeking, finding, preparing, and keeping highly qualified principals and assistant principals. This purpose is also reflected in the AISD *Strategic Plan: 2015–2020*.

Theory of change statement:

> If we create a leadership pipeline that reflects AISD values and a leader development mindset, then we will enable improvements in campus administrator quality with long-term outcomes such as increases in staff retention, student achievement, and campus climate and culture. To do this, the leadership development team develops and implements rigorous selection processes and aligned systems of professional development opportunities.

Figure 2.
Graphical Representation of the Leadership Development Theory of Change



Source. AISD Department of Leadership Development Theory of Change, last revised May 24, 2019

Evaluation of Hiring Process Operations

Two best practices documents based on extant literature (e.g., Arthur & Day, 2010; Campion, Palmer, & Campion, 1997; Taylor & Small, 2002) were developed for the leadership development team. The first of these contains best practices for item development, interviewing, and creating and using rubrics. The second document contains best practices for conducting, scoring, and making decisions related to assessment centers. These documents are in Appendix A. While specific to the needs of the leadership development team, many of the points made in these best practices documents could be generalized to other personnel selection scenarios.

Related to feedback about the operations of the hiring process, several items concerning fairness, scoring, connection between tasks, and the helpfulness of the toolkit from the HireVue and assessment center feedback surveys were analyzed. Looking at fairness, even though most participants reported that they did not know how the assessment center was scored, the majority of both AP and principal participants strongly agreed or agreed that the assessment center was fair. Similar to the assessment center, despite not knowing how the HireVue was scored, the majority of both AP and principal candidates reported that they felt the scoring of the HireVue was fair (Figure 3). This information can help the leadership development team understand the face validity of the new hiring process. Despite being unfamiliar with the scoring of the new process, candidates still trusted that the scoring process was fair. Additionally, APs were more likely than principals to report that the HireVue and assessment center accurately assessed their readiness to enter into the new administrator role (Figure 4). As for the toolkit, of those who used it, 89% of AP candidates found the toolkit moderately to extremely helpful, while 80% of principal candidates found the toolkit moderately to extremely helpful (Figure 5). This information provides the leadership development team with useful insight into perceptions of the toolkit. Qualitative data from the feedback surveys indicated that scoring rubrics, HireVue items, and clear technological instructions may be beneficial, additional tools for preparation. Lastly, most AP and principal candidates drew connections across the different tasks in the assessment center (Figure 6).

Figure 3.

The majority of AP and principal candidates perceived the HireVue and assessment center scoring processes to be fair, despite most not knowing how either were scored.

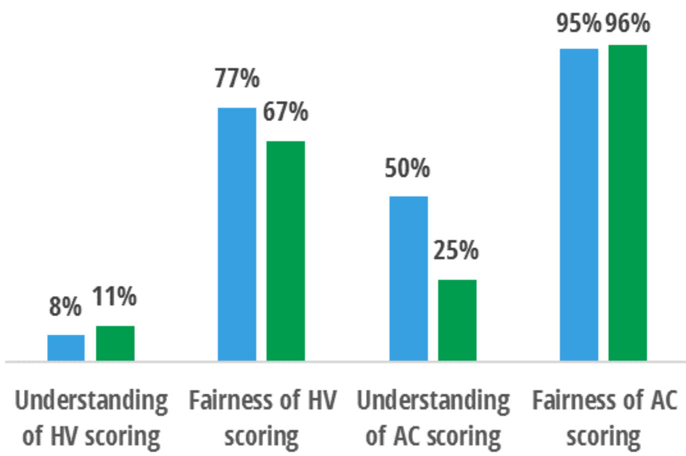
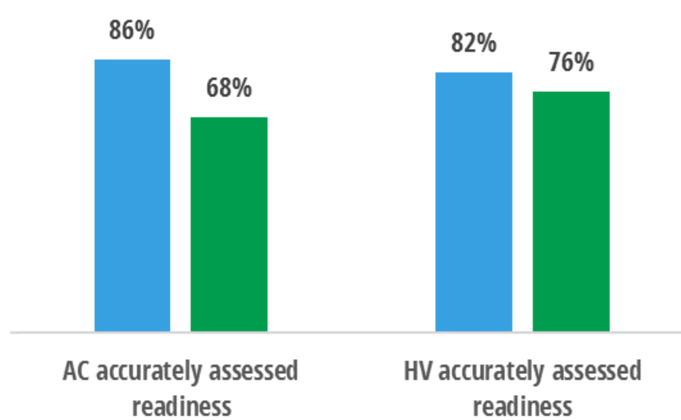


Figure 4.

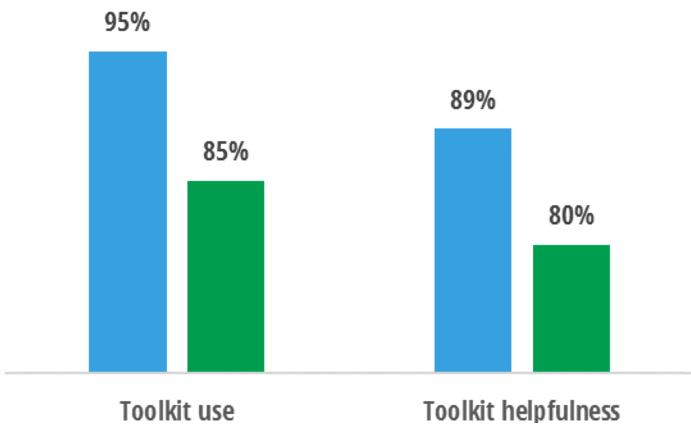
AP candidates felt more strongly than did principal candidates that the HireVue and assessment center assessed their readiness for the new role.



Source. Winter 2019 AP and principal feedback surveys

Figure 5.

AP assessment center candidates used the toolkit more and found it more helpful than did principal assessment center candidates.



Source. Winter 2019 AP and principal feedback surveys

Figure 6.

Fifteen percent or fewer of both AP assessment center candidates and principal assessment center candidates reported they drew no connections across assessment center tasks.

Evaluation of Output and Department Accomplishments

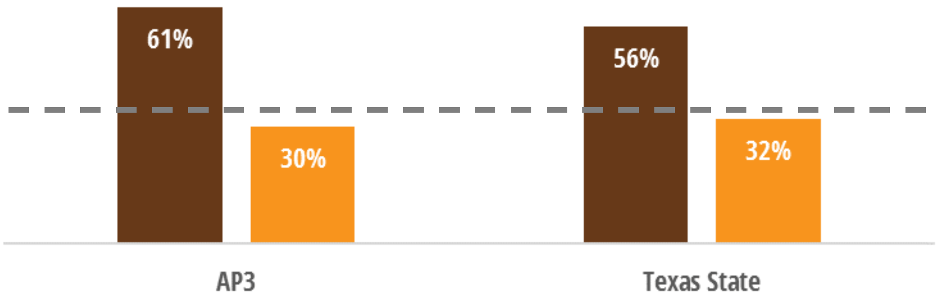
Output is a question of production, not of impact—that is, what evidence indicated that the new hiring process produced what it was supposed to (i.e., a leadership pipeline that produces diverse, equity-focused leaders)? To address the question of production, pass-through rates for APs and principals were calculated and examined using different disaggregations. Additionally, interviews with principals and APs were conducted and summarized to gain further insight into AP candidates' perceptions of the new hiring processes as well as hiring principals' perceptions of accessing and using the highly qualified list.

HireVue Pass-Through Rates: APs

Overall, 33% of applicants who were invited to the HireVue passed their HireVue. Participants in the Texas State program and those in the AP3 program were more likely to pass (at a 56% and a 61% passing rate, respectively) than those who did not participate in the Texas State program or AP3 (Figure 7). These differences in pass-through rates associated with participation in a preparation program suggest alignment between the curricula of these programs and the hiring process. Furthermore, it is understood that the preparation programs are specifically tailored to AISD's leadership value system, or the Campus Administrator Performance Review (CAPR) rubric. This provides evidence more evaluatively that the preparation programs' curricula and the hiring process were all aligned to CAPR, given the stronger performance in HireVue by those who participated in one of these specialized programs in comparison with the performance of those who did not. This trend, in which performance was higher for individuals who participated in a specialized program, is suggested in subsequent sections as well.

Figure 7.

AP participants who were in a preparation program were more likely to pass HireVue than were those not in a preparation program.



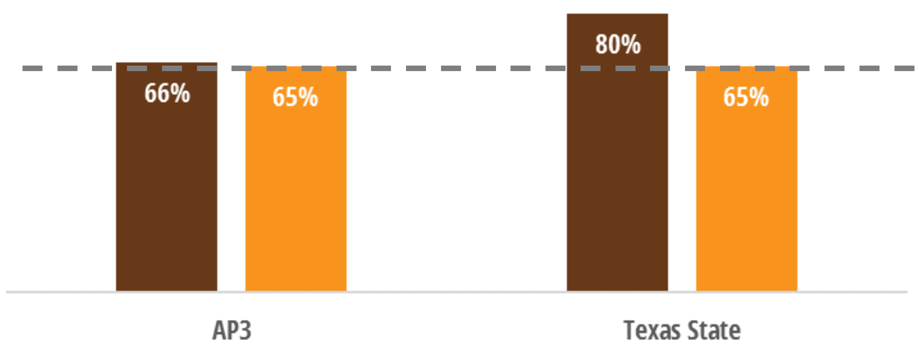
Source. Summer 2018 AP HireVue process data
Note. Dashed line represents percentage of applicants who passed out of those invited to participate (33%).

Assessment Center Pass-Through Rates: APs

Overall, 65% of those who were invited to the assessment center passed. When looking at assessment center pass-through rates, participation in a preparation program appeared to improve passing rates, but less so than was seen for HireVue passing rates. (Figure 8). AP3 cohort members only had a slight advantage over non-members in the assessment center, while Texas State cohort members seemed to have a greater advantage over those who were not in the Texas State program.

Figure 8.

AP participants who were in a preparation program were more likely to pass the assessment center than were those not in a preparation program, with the greatest advantage for those in the Texas State cohort.



Source. Summer 2018 AP assessment center data
Note. Dashed line represents percentage of applicants who passed out of those invited to participate (65%).

HireVue Pass-Through Rates: Principals

Complementary analyses using principal HireVue and assessment center data were conducted to identify if some of the key indicators for successful completion of HireVue

Multiple-Hurdle Approach

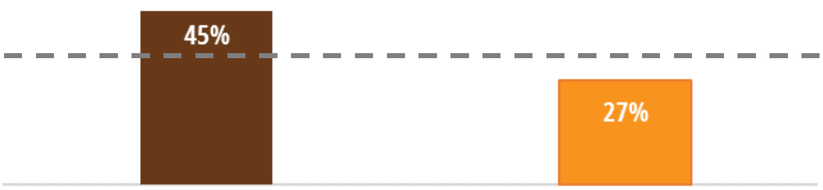
The new AP and principal selection processes use the *multiple-hurdle approach* to hiring. In the multiple -hurdle approach, applicants must pass each selection tool or hurdle in order to continue in the process to the next hurdle. In the new AP and principal hiring processes, applicants must first pass the HireVue interview hurdle in order to participate in the assessment center hurdle. This approach is considered most appropriate when training for the position in question is long, complex, and expensive, when an essential knowledge, skill, or ability for being successful on the job cannot be compensated for by high levels of another knowledge, skill, or ability, and when consequences of making an error in hiring are high (Smith & Angie, 2011). Based on this, the multiple-hurdle approach is an appropriate selection strategy for identifying high quality APs and principals. Moreover, the structuring of a multiple-hurdle approach is also important. Typically, more expensive and time consuming assessments are conducted toward the end of the selection process, while less expensive and less time consuming assessments are toward the beginning. Again, we see that the selection processes used for hiring APs and principals follows this suggestion. The HireVue interviews allow for many applicants to interview for the role very quickly and with less effort, while the assessment center requires much more time and effort.

and the assessment center for principals looked different than what was learned from examining the AP data in detail. Pass-through rates for principal candidates who participated in a HireVue and assessment center in February and March 2019 were examined. These analyses highlight differences and similarities of pass-through rates by participation in P3. In sum, participating in P3 increased passing rates on the HireVue and assessment center.

Only 33% of individuals who were invited to complete a HireVue for a principal role passed their HireVue. Participants in P3 were more likely to pass than those who were not in P3 (Figure 9). Again, this higher passing rate of the HireVue interview for P3 members can likely be attributed to the preparation for an AISD principal role that is provided in the P3 program.

Figure 9.

Principal HireVue participants who were in P3 were more likely to pass HireVue than those who were not in P3.



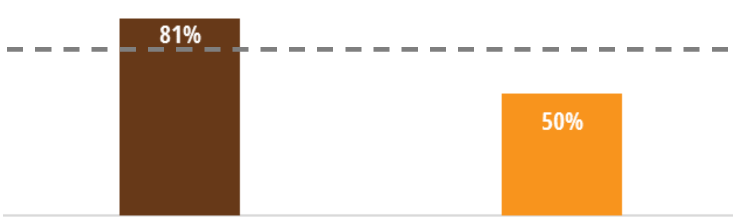
Source. Spring 2019 principal HireVue data
Note. Dashed line represents percentage of applicants who passed out of those invited to participate (33%).

Assessment Center Pass-Through Rates: Principals

Concerning the assessment center, 69% of individuals who were invited to the principal assessment center passed. Overall, we see that those who participated in P3 passed more frequently than did those not in the program. Examining pass-through rates for different groups, 81% of P3 cohort members who were invited to the principal assessment center passed, while only 50% of those not in P3 passed (Figure 10). This provides further evidence that the preparation provided in P3 may help individuals to successfully complete the new hiring processes and gain a spot on the highly qualified list.

Figure 10.

Principal assessment center participants in P3 were more likely to pass the assessment center than were those who were not in P3.



Source. Spring 2019 principal assessment center data
Note. Dashed line represents percentage of applicants who passed out of those invited to participate (69%).

Experiential Data: AP Candidate and Hiring Principal Perspectives

Lastly, to determine how the new hiring process was perceived by those engaged in the new processes, interviews were conducted with four current APs and five current principals. From these interviews, information related to the applicability of the hiring process to on-the-job tasks, the ease of using the new process for hiring principals, and possible points of confusion for both AP candidates and/or hiring principals was gathered. All APs who were interviewed had been hired as a result of successfully completing the HireVue and assessment center, and all principals interviewed had used the highly qualified list of AP candidates (those who had successfully completed the HireVue and assessment center) to hire a new AP on their campus.

Overall, both APs and principals indicated the assessment centers were a worthwhile process that may help to reduce nepotism or a perceived lack of rigor in the district's previous hiring process for campus administration. Additionally, many principals felt the new process allowed them to skip weeding through candidates in Applitrack, and instead pointed them straight to individuals who had been properly vetted. Specifically, principals who had been assessors in the AISD AP assessment centers mostly felt that using the assessment center process for personnel selection was a step in the right direction for the district. Many of these principals also saw alignment between CAPR competencies and standards and the assessment center.

However, some points of confusion or frustration in the hiring process for both APs and principals were discovered through the interviews. Based on anecdotes from APs and principals, we categorized these barriers in the hiring process as (a) education and (b) accessibility/usability.

Education. Principals and APs reported confusion around applying for the pool versus applying for a specific campus-based position. APs and principals seemed mixed on their understanding of the differences between applicants in the pool and applicants on the highly qualified list. Some APs seemed unclear about how the selection process worked after a candidate's successful completion of the assessment center. For example: Do AP candidates who pass the assessment center then need to apply to specific job postings or is applying to the pool and then making it onto the highly qualified list enough? Will and do principals reach out to candidates on the highly qualified list without considering if they have applied to the position on their campus? Are potential APs allowed to visit the principal at schools where they are interested in being employed? Who can or cannot be sent an invitation for an interview? Is the highly qualified list sent out inclusive of everyone that could possibly be hired, or, are campuses sent a subset of candidates selected specifically for that campus?

In conjunction with AP confusion, some principals noted that, as a result of misunderstandings about how AP candidates should apply to positions, candidates would often be invited to interview for an AP position at a school even if the candidate had not applied to that campus's specific job posting. Additionally, in the current system when principals use Applitrack to view the candidate pool or candidates who applied for their campus's position, principals reported being able to see the applications of candidates who were not allowed to be considered because they were not on the highly qualified list. This could create some confusion and frustration on behalf of the hiring principal if some of these other candidates, especially those that appear qualified, cannot be interviewed. Relatedly, some principals felt that being forced to use the highly qualified list limited their autonomy in hiring decisions and that the candidate options on the highly qualified list may not meet the individual needs of or properly fit their campus. Several principals wanted to see more candidates of color and more bilingual candidates due to campus needs. Lastly, several principals voiced frustration about candidates on the highly qualified list declining to interview.

Accessibility/Usability. Due to the current system, principals' ability to understand, access, and use the highly qualified list to identify the best candidates for their campus was hindered. As it stands, principals who have an AP vacancy are emailed a static version of the highly qualified list. The static nature of the list makes using it more difficult. Some principals mentioned reaching out to candidates who had already been hired elsewhere, which they felt was a misuse of their time. This issue appeared to happen to some hiring principals before reaching out to candidates to schedule an interview and between the candidates' interview(s) and an offer extension. Moreover, to access other information about the candidates contained on the highly qualified list, principals must log in to Applitrack and search for the candidate from the list in the applicant pool. Candidates' resumes, letters of interest, and references are all housed in Applitrack, and therefore the use of Applitrack is still necessary for hiring. Additionally, principals must cross-reference the highly qualified list with those who apply to their campus's position opening and with those who are in the applicant pool, to ensure that only candidates on the highly qualified list are selected for interviews. These steps for locating candidates in Applitrack and cross-referencing Applitrack with the highly qualified list hinders principals' ability to efficiently access information about candidates who have applied for and are qualified to work on their campus.

Summary and Recommendations

Summary of Evaluation Findings

The implementation of the new hiring process began in early summer 2018, and assessment centers have been consistently conducted since. As originally planned, assessment centers for APs were to be implemented in 2018–2019 and would be used to inform the future implementation of assessment centers for principal hiring. However, both AP and principal assessment centers were implemented in 2018–2019. Therefore, due to this change in workflow with the addition of the new hiring process for principals, we worked with the leadership development team to conceptualize the work, goals, and areas for impact of the department. This resulted in the development of the ToC and logic model, which helped to highlight the wide range of work the department is involved in, the goal the department hopes to achieve through engaging in that work, and the ultimate impacts the department aims to achieve if that goal is met.

To help ensure that the new hiring processes were operating as designed, a workshop was held for the leadership development team to provide information on best practices for conducting assessment centers and interviews. Additionally, survey feedback from APs and principals helped clarify perceptions of the hiring process. Data related to the usefulness of the toolkit; fairness of the process; understanding of the scoring; and relevance of the items or tasks to knowledge, skills, abilities, and other characteristics (KSAOs) necessary to be successful in an AP or principal position were gathered, and results indicated that applicants had generally positive perceptions of the new hiring processes.

Preliminary data guided our understanding about whether the new hiring processes were producing what they were intended to produce: a group of highly qualified, diverse, equity-focused leaders. While pass-through rates regarding demographic diversity were not included in this report, this data should be further explored as sample sizes become larger, as current data likely do not provide a full picture of the potential of adverse impact in either the HireVue or assessment center processes. Pass -through rates did indicate that participating in a preparation program (i.e., either AP3, P3, or the Texas State program) helped participants pass both the HireVue and the assessment center, and therefore increased the odds of those participants gaining a spot on the highly qualified list. While data regarding scores on individual items or tasks was not displayed here to maintain their integrity, evidence also suggests that scores on both the HireVue and assessment center were generally normally distributed, showing that the items and tasks can discriminate low and high performance. Analyzing mean scores of AP candidates on HireVue items and assessment center tasks, while not displayed in this report, also showed that items and tasks were of similar difficulty, with a slight positive skew on the HireVue item scores and a slight negative skew on the assessment center task scores. Lastly, interviews with both APs and hiring principals provided evidence that the new hiring process was perceived as worthwhile, with the recognition that the process is new and will be further refined in the years to come. Feedback from the interviews (e.g., the need to allow principals autonomy in hiring decisions, the need to increase diversity on the highly qualified list, and the need to clarify the process for APs after they have successfully completed the assessment center) will help to inform changes to the hiring process for subsequent years of implementation.

Recommendations and Future Directions

As data analyses and feedback were available, the leadership development team was provided formative recommendations that they were able to implement throughout the year. For example, the number of days an applicant has to complete their HireVue was increased based on survey feedback, and scoring procedures were adjusted based on the workshop hosted in the fall. In the upcoming years of HireVue and assessment center implementation, several more recommendations based on the departments' workflow and the summative results provided in this report are to:

- Continue to monitor pass-through rates and examine if adverse impact has occurred as more candidates participate in the hiring process. For future AP and principal openings, the district should engage in more strategic recruitment to increase applications from minority candidates, with special attention to recruitment of male, African American, and Hispanic APs and principals.
    - Strategic recruitment should also be used when selecting cohort members for AP3 and P3. Given the advantage that the curricula these programs provide in regard to passing HireVue and the assessment center, selecting a diverse group of individuals for those cohorts could help to bolster the diversity of those who make it on the AP and principal highly qualified lists.
- Move toward a standard practice for determining passing and failing scores on the HireVue and assessment center, based on suggestions from the workshop and best practices documents.
- Create clear messaging to and education for both principals and APs about the logistics of the hiring process to help curb confusion and educate stakeholders about how the process is designed to operate. For example, messaging should highlight that the new hiring process is designed to increase diversity, given that assessment centers are empirically found to not produce biases in selection, and to help streamline administrative hiring (Arthur & Day, 2011). Additionally, it would be beneficial for highly qualified AP candidates to be urged to apply for jobs at individual campuses and not just wait for principals to reach out to them. APs should also be encouraged to only apply to campuses where they can realistically see themselves accepting a position. There will likely be fewer instances of candidates declining to interview or declining a job offer if principals are only inviting candidates who have applied to their campus's job opening. Principals also need to be educated more about the new hiring process for APs. Clear instructions to only interview candidates on the highly qualified list, and specifically to start their candidate search with those who have applied for their campus's open position would be beneficial.
- Consider if the highly qualified lists can be shared with necessary staff in a dynamic way to increase the accessibility and usability of the lists. This way, principals can see when someone from the list is no longer available because that person chose to drop off the highly qualified list or was hired elsewhere. Additionally, if the live list

Adverse Impact Explanation

Adverse impact, as assessed by the four-fifths rule, specifies that if the selection rate for any group is less than 80% of the selection rate for the group with the highest selection rate, there is statistical evidence of adverse impact. This newly calculated ratio is called an *impact ratio*. In the case of small sample sizes, literature cautions against calculating adverse impact as a means to identify bias, since selection bias is often overidentified when sample sizes are too small (Collins & Morris, 2008; Roth, Bobko, & Switzer, 2006). Specifically, adverse impact should only be calculated using groups that make up more than 2% of the applicants (Uniform Guidelines, U.S. Equal Employment Opportunity Commission, 1978).

An example of calculating adverse impact is as follows:

Gender	Invited	Hired	Selection Rate
Male	50	30	30/50= .60
Female	30	10	10/30= .30

To calculate the impact ratio, we would take the selection rate of the minority group (in this case, women; .30) and divide it by the selection rate of the majority group (in this case, men; .60).

Therefore, we would calculate that .30/,60= .50, or 50%. Since 50% is less than 80%, we would conclude that there is statistical evidence of adverse impact against women in this example.

- were hosted as a tile in the portal, there would be potential for data from Applitrack to be piped into the tile, making the tile a one-stop shop for all information related to highly qualified candidates. The tile could also hold information related to whether candidates have already been pulled for an interview, and if so, where they have been pulled for an interview. This could help hiring principals see how the list is changing in real time, and also could serve as another data point for principals to use when assessing a candidate's fit for their campus. For example, if a hiring principal sees that a similar campus has interviewed a candidate, then that candidate may also be a good fit for the hiring principal's campus. Other possibilities for the tile include the options for communication between hiring principals and candidates (e.g., send interview requests, accept interview offers) and the capability for the list to automatically remove candidates as they have been hired. Flags for other characteristics, such as participation in a preparation program or bilingual status could also be added for easier candidate searching.
- Implement a standardized process for archiving HireVue and assessment center applicants. The creation of an Excel or Google sheet to serve as a database that contains all AP and principal applicants who have ever completed a HireVue or an assessment center would aid in efficiency and the ability to analyze applicant data.
- Streamline processes that occur before and after the HireVue and assessment center.
    - Determine if the AP and principal pools could always be open for applications. Relatedly, determine if the HireVue interviews could be scored on a rolling basis. That is, if applications for AP and principal positions are coming in steadily, could applications be reviewed once a week, with invitations for HireVue extended to applicants weekly? Positions being consistently open for applicants and HireVue invitations being extended more regularly could (a) allow for a larger window of time for applicants to complete their HireVue (given no impending deadlines); (b) lessen the number of applicants who need to be reviewed at a single time when determining who should/should not receive a HireVue invitation; (c) result in more applicants applying to the positions (given that the posting is always open), which would also likely result in more qualified candidates applying to the positions; (d) spread out the number of HireVue interviews that need to be scored by the leadership development team at a given time; and (e) engage AP and principal applicants in the hiring process more quickly. This new process of reviewing applicants in Applitrack and extending HireVue invitations weekly, as well as scoring HireVues on a rolling basis, should create greater process efficiency for the leadership development team and should have a positive impact on applicants, given that they will receive their HireVue invitations shortly after submitting an application.
    - Determine ways to ensure both candidates and hiring staff are aware of the appropriate actions that need to be taken after a candidate is added to the highly qualified list.

After progress is made on ensuring that all pieces of work are being implemented and are performing well, the department's current and desired levels of performance toward the department's goals can be measured. After the new processes have been in place for more time, it will become increasingly appropriate to determine if the goal of the department is being met. Moreover, after the department's work has been fully implemented, and performance is measured around its goals, then we will focus on evaluating potential impacts.

References

Arthur, W., Jr., & Day, E. A. (2011). Assessment centers. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology, Volume 2: Selecting and developing members for the organization* (pp. 205–235). Washington DC: American Psychological Association.

Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology*, *50*(3), 655–702.

Collins, M. W., & Morris, S. B. (2008). Testing for adverse impact when sample size is small. *Journal of Applied Psychology*, *93*(2), 463–471.

Connell, J. P., & Kubisch, A. C. (1998). Applying a theory of change approach to the evaluation of comprehensive community initiatives: progress, prospects, and problems. *New approaches to evaluating community initiatives*, *2*(15-44), 1-16.

Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, and Department of Justice. (1978). Uniform guidelines on employee selection procedures. *Federal Register*, *43*, 38290–38315.

Roth, P. L., Bobko, P., & Switzer, F. S. (2006). Modeling the behavior of the 4/5ths rule for determining adverse impact: Reasons for caution. *Journal of Applied Psychology, 91*, 507–522.

Smith, C., & Angie, A. D. (2011). *Developing a Multiple Hurdle Approach to Hiring*. Paper presented at the 2011 annual conference of the International Personnel Assessment Council, Washington, D.C.

Taylor, P. J., & Small, B. (2002). Asking applicants what they would do versus what they did do: A meta-analytic comparison of situational and past behaviour employment interview questions. *Journal of Occupational and Organizational Psychology*, *75*(3), 277–294.

W. K. Kellogg Foundation. (2004). Introduction to Logic Models. In *Logic Model Development Guide.* (pp. 1-12). Battle Creek, Michigan.

AUSTIN INDEPENDENT SCHOOL DISTRICT

Paige DeBaylo, Ph.D.

Department of Research and Evaluation

1111 West 6th Street, Suite D-350 | Austin, TX 78703-5338
512.414.1724 | fax: 512.414.1707
www.austinisd.org/dre | Twitter: @AISD_DRE

August 2019

Publication 18.25

Appendix A.

Best Practices and Heuristics: Assessment Centers

Use the questions listed in this document to examine your assessment center processes and exercises. These questions are included here to get you thinking about some of the most important factors for creating, scoring, and implementing successful assessment centers for personnel selection.

Are your tasks measuring what is important? How do you know if your system of scoring is measuring what is important?

Assessment center tasks are designed to capture information related to underlying constructs/competencies, or they are direct representations of activities the candidate would perform on the job, or they are both. Therefore, each exercise within the assessment center should represent at least one leadership framework competency and/or the assessment center tasks should mimic an on-the-job task. Some questions to ask yourself:

- Are your exercises representative of on-the-job tasks performed? Or are they activities that allow candidates to demonstrate competencies?
- For example, is the inbox task designed to determine how well the candidate can send emails? Or is the inbox task designed to allow the candidate to demonstrate competencies in the leadership framework?
- If the exercises represent job tasks, then does the score on an exercise provide a measure of job skill?
- For example, if the inbox task is designed to represent emailing, does the score provided on the inbox task give us an idea of a candidate's emailing skill?
- If the activities allow candidates to demonstrate competencies, then do their scores provide measures of relevant competencies?
- For example, if the inbox task is designed to represent leadership framework competencies, does the score provided on the inbox task give us an idea of a candidate's competence in strategic thinking or instructional practice?
- If activities are both job relevant and reflective of competencies, then do you have systematic scoring in place for both tasks and competencies?
- For example, if the inbox task is designed to represent both the capacity to send email and leadership framework competencies, are candidates systematically scored on both components within the inbox task?

How are you scoring assessment center tasks?

It is recommended to always use detailed, anchored rating scales, which provide context to the numerical ratings that are assigned to task or interview responses. An answer that you deem as a 3 may be very different from the type of answer that someone else views as a 3, and therefore providing as much context for what type of candidate answer constitutes each numerical rating helps to standardize the scoring. By using standardized scoring metrics, there is a decrease in the potential for bias and an increase in the likelihood that scores in the assessment center will be predictive of job performance and other important outcomes. Some questions to ask yourself:

- Are you using anchored rating scales?
    - If yes, do you have descriptions of what the different ratings mean?
    - If yes, do you have examples of responses that match the different ratings?

Below is an example of a rating scale that contains both a description and a space for an example response for each rating 1 through 4 for a given competency. Use past performances on the assessment center to fill in the "Example response" where necessary/applicable. Note that currently all behaviors/look-fors are listed in each rating. Prioritizing the behaviors/look-fors may help refine the descriptions for each rating. For example, maybe understanding diversity is more important than understanding social emotional learning, or maybe training someone to understand social emotional learning is easier than training someone to understand diversity. Therefore, maybe it is vital for "understanding diversity" to be in the 4 rating, but less vital for "understanding social emotional learning" to be in the 4 rating.

Competency	Rating			
	1	2	3	4
Child-centered lead-ership	Lacks an understanding of diversity, inclusivity, social emotional learning, and equity and access for all students. *Would need substantial training in this area.*	Demonstrates some understanding and ability to provide diversity, inclusivity, social emotional learning, or equity and access for all students. *Some training necessary in this area.*	Provides an understanding of supporting diversity, inclusivity, social emotional learning, or equity and access for all students. *Would need no training in this area.*	Provides a strong understanding of and many experiences related to diversity, inclusivity, social emotional learning, and equity and access for all students. *Could mentor and teach others in this area.*
	Example response:	Example response:	Example response:	Example response:

Additionally, using more than one rater is recommended for scoring assessment center tasks when possible. This reduces the likelihood of bias and often provides for richer information on the candidate. Ask yourself the following questions:

- Do you have multiple raters for each task?
    - If yes, do those raters provide individual ratings?
- Are those individual ratings being averaged at the item level?

Are you using weighting for scores?

Weighting of exercises or competencies should be properly considered and assigned. Here are some factors to consider in weighting:

- Relative importance—deciding which selection exercises or competencies contribute the most to successful job performance. The more important a competency is, the higher the exercises that represent that competency should be weighted.
- Consequence of failure—deciding how damaging not being competent in a specific area might be to the candidate's future job performance. The more damaging not having a competency is, the higher the exercises that represent that competency should be weighted.
- Time to achieve competence—deciding which competencies take longer for an employee to master once they are on the job. The longer a competency takes to

master, the higher the exercises that represent that competency should be weighted (given that training is expensive).

If you are not using weighting:

- Are you assuming that what you are measuring is equally important for job performance, equally damaging to job performance if not present, and equally difficult to train an employee on once on the job?

If you are using weighting:

- Are your highest weights given to things that are most important for job performance, most damaging to job performance if not present, and most difficult to train an employee on once on the job
- Are your lowest weights given to things that are least important for job performance, least damaging to job performance if not present, and least difficult to train an employee on once on the job?

Current Relationships Between Competencies and Exercises

The following matrix depicts how competencies are currently mapped to exercises:

Competency	**Exercises**			
	Data (35%)	Observation (20%)	Inbox (10%)	Interview (35%)
Instructional leadership	✓	✓	✓	✓
Talent development	✓		✓	
Child-centered leadership	✓		✓	✓
Culture building	✓			✓
Strategic thinking			✓	✓
Executive leadership	✓		✓	✓

Note. Matrix is current as of September 21, 2018. Changes have been made to the process since that may impact the alignment between competencies and exercises.

How are you using your scoring data?

There are a variety of ways to use, display, and transmit candidate scores. The following questions will walk you through questions to ask yourself when considering how to best use scoring data to make the most informed personnel decisions.

- Are you reporting out on performance?
    - If yes, what are you reporting on?
- Are the data points summarized in a report capturing what you care about?
    - For example, are the data points related to job tasks or job competencies?
    - If yes, who is receiving the report?
- Are data collected for the purpose of getting candidates to the hiring pool and/or for the purpose of providing hiring managers with relevant information on the candidates?
    - If for the purpose of providing hiring managers with relevant information, is

candidate performance information for tasks and/or competencies included for review by those making selection decisions?

    - If no, how is candidate assessment center performance information being provided to those making selection decisions?

- Do candidates receive performance information for tasks and/or competencies?

There are also a variety of ways to determine who did well on the assessment center and who did not. Distinguishing high achievers from lower achievers will be important when deciding who makes it into the applicant pool. Following are some questions to consider when determining high and low performance as well as who makes it into the applicant pool.

- Are you using numbers/scores to determine passing the assessment center?
    - If yes, where does the number you use to determine passing come from?
        - Number of openings (e.g., there are 100 openings, so we need to pass a lot of people)
        - Minimum qualifications/standard (e.g., the minimum scores on competencies/exercises determine passing)
        - Top performers by percentile (e.g., the top 20% of performers in each assessment center make it to the pool)
        - Assessment of pass rates/distribution of scores (e.g., use the distribution of the scores to determine the cutoff score for passing)
        - Consistent criterion (e.g., use the same cutoff score for all assessment centers)

You may also be using the scores on the assessment center to determine how your tasks/exercises are functioning. It is important to be aware of tasks that might be too hard for candidates, too hard for assessors to rate, or not be representative of what you think. Here are some questions to keep your tasks in check.

- Are you using candidates' scores to evaluate the tasks themselves?
- Are you aware of difficulties for assessors on any particular task?
    - For example, do assessors mention that a task is hard to score, do assessors mention that it is hard to come up with a rating scale for a task, or do you see evidence of poor interrater reliability?
- Are you aware of difficulties for candidates on any particular task?
    - For example, do scores show that all candidates pass certain tasks or items within a task or that all candidates fail certain tasks or items within a task?

Assessment center ratings: the rating given to each exercise and/or competency within an exercise. Each exercise should have its own rating and/or each competency should have its own rating (averaged across all exercises). Each candidate therefore has a rating for each assessment center exercise and/or each leadership framework competency (possible 10 ratings in total). Use the following table as a reference when examining ratings.

Checklist	Evaluation	Comments
Appropriate weights assigned to exercises	✓ ✗	
Appropriate weights assigned to competencies	✓ ✗	
Individual ratings by exercise included in candidate report	✓ ✗	
Individual ratings by competency included in candidate report	✓ ✗	
Individual ratings by exercise included in candidate view in applicant pool	✓ ✗	
Individual ratings by competency included in candidate view in applicant pool	✓ ✗	

Assessment center development and training: a common process to go through to ensure the assessment center captures job-relevant behaviors and competencies and to ensure that assessors understand how to rate exercises. Each of the four exercises

Checklist	Evaluation	Comments
Determine the purpose of conducting the assessment center	✓ ✗	
Define competencies and descriptions of exercises	✓ ✗	
Assessors given examples of behaviors that represent each competency	✓ ✗	
Assessors familiarize themselves with behaviors, competencies, and exercises	✓ ✗	
Assessors determine which behaviors fit which competencies and exercises	✓ ✗	
Creation of anchored rating scales with descriptions and examples for each exercise	✓ ✗	
Assessors practice observing the behaviors (e.g., via video, role play) in practice assessment	✓ ✗	
Assessors categorize behaviors into competencies in practice assessment	✓ ✗	
Assessors assign ratings for each exercise in practice assessment using rating scale	✓ ✗	
Assessors discuss observations and ratings with other assessors to build common frame of reference	✓ ✗	

should have behaviors (i.e., look-fors) that are mapped onto an anchored rating scale, with descriptions and an example response. Use the following table as a step-by-step guide for assessment center development and assessor training.

Best Practices and Heuristics: Interviews

Job analysis: a lengthy process to assess relevant tasks an employee does on the job as well as personal characteristics an employee needs to be successful on the job. A possible way around doing an actual job analysis, or at least a starting point, could be using O*NET: https://www.onetonline.org/link/summary/11-9032.00

Types of interviews: there are structured, unstructured, and semi-structured interviews within personnel selection processes. Structured interviews are based on a job analysis, consist of standard ways of gaining job relevant information, are quantitatively evaluated, and focus on job criteria. Conversely, unstructured interviews are based on the interviewer's personal beliefs and intuition, consist of unstandardized ways of gaining information, and can be biased by impressions. The recommendation is to use structured interviews.

Types and examples of structured interview questions: different types of interview questions will provide different types of answers. Some structured interview question types and examples are:

- Situational questions (SQs): ask candidates what they would do in a given, future situation
    - Suppose a coworker was not following standard work procedures. The coworker was more experienced than you and claimed the new procedure was better. Would you use the new procedure?
    - Suppose you were giving a presentation, and a difficult technical question arose that you could not answer. What would you do?
- Past behavior questions (PBQs): ask candidates about past work experiences and behaviors
    - Based on your past work experience, what is the most significant action you have ever taken to help out a coworker?
    - Can you provide an example of a specific instance when you successfully engaged families and the community to promote student success?
- Background questions: ask candidates more general background information
    - What work experiences or training do you have for working in this environment?
    - What experience have you had with developing systems of curriculum, instruction, and assessment?
- Job knowledge questions: ask candidates specific questions related to job tasks
    - What steps would you follow to conduct a brainstorming session with a group of employees on professional development opportunities?
    - What factors would you consider when cultivating a child-centered school community?

General Tips for Item Writing:

- Use straightforward and simple language
- Write items that candidates are likely to answer differently

- Avoid double-barreled items (e.g., how satisfied are you with your pay and working environment?)
- Word items carefully to ensure they are measuring the competency you intend them to measure (e.g., executive leadership items need to be discerned from child-centered leadership items)

PBQ item writing: items should be written to gain details about a candidate's past experiences that reflect identified job-related competencies. A common framework used for developing a structured, behavioral interview question is the STAR model. This is outlined as follows:

- **S**ituation: what situation was the candidate in?
    - "Tell me about a time…"
- **T**ask: what task did the candidate need to accomplish?
    - "Were you were faced with multiple competing deadlines?"
- **A**ction: what actions did the candidate take to accomplish this task?
    - "What did you do and…"
- **R**esults: what were the results of these actions?
    - "How did it turn out?"

So your completed PBQ would read, "Tell me about a time when you were faced with multiple competing deadlines. What did you do, and how did it turn out?"

Rating answers: use descriptively anchored rating scales, like the one that currently exists for the assessment center interview. However, providing an example answer as well as descriptions of what the answer should entail may help rating. If it is hard to generate an example/description of what the answer should be, maybe the question is too difficult.

A typical scale template* used to rate competencies within an employment interview is as follows:

5 = Superior skills in this competency; could mentor or teach others
4 = Good skills in this competency; above-average ability is apparent
3 = Adequate skills in this competency; no additional training is needed at this time
2 = Marginal skills in this competency; some training would be required to bring skills up to an acceptable standard
1 = Not competent in this area; competency needs substantial development

* This could be translated to a 4-point scale to match current ratings for HireVue and the assessment center interview/exercises.

The following table shows an example of a rating scale that contains both a description and a space for an example response for each rating 1 through 4 for a given competency.

Use past performances on the assessment center to fill in the "Example response" where necessary. Note that currently all behaviors/look-fors for an item/competency are listed in each rating. Prioritizing the behaviors/look-fors may help refine the descriptions for each rating. For example, maybe understanding diversity is more important than understanding social emotional learning, or maybe training someone to understand social emotional learning is easier than training someone to understand diversity. Therefore, maybe it is vital for "understanding diversity" to be in the 4 rating, but less vital for "understanding social emotional learning" to be in the 4 rating.

Competency	Rating			
	1	2	3	4
Child-centered leadership	Lacks an understanding of diversity, inclusivity, social emotional learning, and equity and access for all students. *Would need substantial training in this area.*	Demonstrates some understanding and ability to provide diversity, inclusivity, social emotional learning, or equity and access for all students. *Some training necessary in this area.*	Provides an understanding of supporting diversity, inclusivity, social emotional learning, or equity and access for all students. *Would need no training in this area.*	Provides a strong understanding of and many experiences related to diversity, inclusivity, social emotional learning, and equity and access for all students. *Could mentor and teach others in this area.*
	Example response:	Example response:	Example response:	Example response:

Adverse impact: bias that occurs as a result of a selection process (e.g., interview, AC, personality tests). The Equal Employment Opportunity Commission (EEOC) typically determines adverse impact using the four-fifths (or 80%) rule. This rule mandates that a selection rate for any race, sex, or ethnic group that is less than four-fifths of the rate for the group with the highest rate will generally be regarded as evidence of adverse impact. Essentially, you can divide the hire rate for the group of interest by the group with the highest hire rate. If this number is less than 80%, there is evidence of adverse impact. Some ways to lessen adverse impact are as follows:

- Use structured interviews in comparison with unstructured interviews
- Rate the applicant on an item after each item, as opposed to at the end of the interview
- Use multiple interviewers
- Ensure all information obtained through the selection process is job related

Example:

	Applicants	Hired	Selection rate
Male	80	48	48/80 = .60, **60%**
Female	40	12	12/40 = .30, **30%**
Impact ratio: .30/.60 = .50, **50%**			

Here, the impact ratio is less than .80, or 80%, which provides evidence, based on the four-fifths rule, that there is adverse impact.

Fifteen components of structured interviews: research suggests all interviews should be conducted according to a set structure. Fifteen components of structured interviews have been proposed to strengthen the validity of the interview process. These components are explained in detail and also have been incorporated into a checklist. Feel free to use this checklist to assess the quality of the interviews you conduct.

1. Base questions on a job analysis

To determine that all questions are soliciting information that is job relevant, and to determine that we know what *job relevant* means, conduct a job analysis prior to developing interview items. Doing this will ensure you receive answers that provide information that is applicable to the job and will decrease chances for bias.

2. Ask each candidate the exact same questions
3. Limit prompting, follow-up questions and requests for elaboration
4. Control use of supplemental information
5. Do not allow questions from the candidate until after the interview

These components, if violated, all increase chances of bias. By not assessing candidates on the same information or not providing candidates with the same opportunities to provide information, interviews become less structured and more susceptible to bias. Additionally, not assessing candidates on information that is applicable to job tasks introduces the opportunity for candidates to be assessed on irrelevant information. The linkage between interview performance and job performance is strongest when interviews assess only job-relevant information.

6. *Use better types of questions*
7. *Use long interviews/more questions*

Better types of questions are either situational questions (SQs) or past behavior questions (PBQs), with PBQs being slightly more favorable. By asking questions in PBQ format, applicants are forced, by the nature of the question, to provide concrete examples of instances in which they have shown competence in the area being assessed. While SQs also force applicants to give an example, the example provided is how they plan to respond to a certain scenario and could be easily faked. PBQs also may be more preferable because they follow the notion that past behavior predicts future behavior. Additionally, questions about attitudes, goals, and self-evaluation are less structured, may not produce quality answers, and are poorly linked to job performance. Interviews should also be long enough to allow time for enough information to be obtained from the candidate. The more job-relevant information that can be gained from the candidate during the interview, the more likely we are to understand the candidate's potential. Also, consider the appropriateness of the question content in relation to the candidate's background and the position's needs. For example, it may be more appropriate to ask teachers about teacher leadership experiences that transfer to an AP role as opposed to asking about school leadership experiences. A teacher may not have a lot of school leadership experiences but is likely have teacher leadership experiences that may be applicable for an AP position.

8. *Rate each answer and use multiple scales*
9. *Use detailed anchored rating scales*
10. *Use multiple interviewers*
11. *Use the same interviewers across all candidates*

Each item within the interview should have its own rating, as opposed to giving one rating to the overall interview performance. This allows us to better understand a candidate's performance on the interview. Also, when possible, multiple raters should provide ratings for each item within the interview. Additionally, each rater should provide their rating for an item within the interview based on a detailed anchored rating scale. Providing example answers as well as descriptions of what an answer at each rating level (1 through 4) should entail within the rating scale will structure the rating process. If it is hard to generate an example/description of what the answer should be, maybe the question is too difficult. Having the rating scale will standardize the ratings across raters and bring more structure to the interview, thus reducing chances of bias and chances of a weak link between interview performance and job performance.

12. Take detailed notes

It is best to take extensive notes during an interview (after a candidate responds to each item) based on facts and the answers provided. Notes taken after an interview and notes taken that contain information about evaluations or judgments of the candidate may not be as helpful. Notes taken after an interview may also miss important information from an item that was forgotten, and notes taken that contain evaluations or judgments of the candidate are likely to contain job-irrelevant information. Additionally, it is best to take notes that are as detailed as possible to provide yourself with as much job-relevant information as possible to inform your interview ratings.

13. Do not discuss candidates/answers between interviews

Discussing a candidate or his or her answers between interviews provides the opportunity for other interviewers'/raters' perceptions of the candidate to change. Discussions of the candidate can cause irrelevant information to enter the evaluation process and can change the standards between interviews. This is something to be extra aware of for panel interviews.

14. *Provide training to interviewers*

All interviewers should receive training on interviewing. This training should include practice interviews and practice rating interviews and exposure to and practice with using the different detailed anchored ratings scales (i.e., rubrics). Interviewers should also be made aware of the best practices checklist during the interview process. This will aid interviewers' ability to conduct structured interviews that assess only job-relevant information.

15. *Use statistical rather than clinical prediction*

Use weighting procedures that are decided beforehand to weigh items by importance within an interview, as needed. Ratings for items should be based on something (a rubric/anchored weighting scale) and not just on intuition or judgment. Thoughtful weighting should be performed to score candidates as opposed to assigning weight to candidate's scores based on judgment or feelings toward a candidate. To ensure weights are thoughtful, higher weights should be given to criteria that have the strongest relevance to the job. The following factors are important to consider if weighting items:

- Relative importance: deciding which selection exercises or competencies contribute the most to successful job performance. The more important a competency is, the higher the item that represents that competency should be weighted.
- Consequence of failure: deciding how damaging not being competent in a specific area might be to the candidate's future job performance. The more damaging not having a competency is, the higher the item that represents that competency should be weighted.
- Time to achieve competence: deciding which competencies take longer for an

employee to master once they are on the job. The longer a competency takes to master, the higher the item that represents that competency should be weighted, given that training is expensive.

Interview Best Practices Checklist

Checklist	Evaluation	Comments
Item quality and development		
• Were my questions based on a job analysis?	✓ ✗	
• Did I ask this candidate the exact same questions we ask other candidates?	✓ ✗	
• Did I use high-quality questions, such as past behavior or situational questions?	✓ ✗	
• Were the number of questions/length of the interview adequate to capture enough information about the candidate?	✓ ✗	
Interview administration		
• Did I limit the amount of prompting, follow-up questioning, and asking for elaboration?	✓ ✗	
• Did I let other information about the candidate (e.g., education, recommendations) have an impact on my interviewing or perceptions of the candidate?	✓ ✗	
• Did I wait until after the interview to allow for questions from the candidate?	✓ ✗	
• Did I rate each interview question separately and during the interview?	✓ ✗	
• Did I use a detailed anchored rating scale?	✓ ✗	
• Did I take notes during the interview that were based on facts and candidate answers?	✓ ✗	
• Are these interview questions properly weighted to reflect importance of the items?	✓ ✗	
Interviewer processes		
• Were there multiple interviewers for this candidate?	✓ ✗	
• Were the interviewers for this candidate people who commonly conduct our interviews (i.e., were they trained on appropriate rubrics)?	✓ ✗	
• Was I involved in a discussion about this candidate with my colleagues before final interview ratings were provided?	✓ ✗	
• Did I receive adequate training (e.g., exposure to and practice with the rubric, practice interviewing) before conducting this interview?	✓ ✗	