When Examinees Cannot Test: The Pandemic’s Assault on Certification and Licensure

Michael G. Jodoin and Jonathan D. Rubright
National Board of Medical Examiners

As outlined in the Special Section Introduction (this issue), the COVID-19 pandemic wreaked havoc on the world economy as shelter-in-place regulations forced individuals to stay at and work from home. Brick and mortar testing centers, whether run as part of a for-profit company, a cooperating educational facility with a history of paper and pencil administrations, or centers established for performance assessments, were largely forced to close to comply both domestically and internationally. As a result, assessment organizations who rely on vendors or partners to provide physical space saw their primary, and in many cases, sole test delivery channel collapse overnight. This short, invited manuscript focuses on the implications for certification and licensure assessment organizations as a result of these widespread disruptions. Although we touch briefly on issues that will resonate for any assessment organization, we have attempted to share more details on issues that are more likely, challenging, and specific to the certification and licensure space.

In the specific context of our organization and other organizations in the health care space, demand for health care providers has soared, setting up a paradox whereby the same force that amplified the need for essential workers stilled the educational, certification, and licensure pathways responsible for introducing new providers into the work force. Certification and licensure organizations were left in the uncomfortable position of needing to advocate for the continued value of testing in upholding minimal competency to protect the health of the public, while also being unable to deliver that value to either the examinees needing to test nor the public requiring the services of health providers in greater numbers. This was complicated by a multitude of secondary uses, including progression within educational programs, graduation requirements for students, accreditation for institutions that provide educational programs, and selection decisions for specialized training. The impact of this pandemic on examinees, educational institutions, regulatory bodies, and assessment organizations has been immediate, yet the changes, whether correctly or incorrectly assumed to be temporary or permanent, will reverberate into the future of testing in the professions.

Impact on Stakeholders
The first stakeholder immediately impacted by pandemic regulations were examinees who awoke to phone calls and emails from test centers that their scheduled exams were canceled as test centers closed. And these were among, in some sense, the “lucky” examinees: many test takers had their testing events stopped partway through their testing experience as select centers closed part way through their normal operating hours. In response to these interruptions, our organization executed preexisting processes normally executed in response to small-scale issues (e.g., a power outage at a specific site) that use statistical models to ensure appropriate pass-fail decisions are honored, thus insuring both public protection responsibilities while minimizing impact on examinees (Feinberg, 2020). For licensure examinations that are often very long to minimize misclassification errors and for students testing under accommodations where exams might span multiple days, these approaches had significant positive impact compared to policies and approaches that count all unadministered items as incorrect and require score expungement and retakes that may not be possible.

As uncomfortable as cancelations were, the uncertainty surrounding when testing would resume and when a rescheduled slot would be next available was arguably worse. Test takers requiring scores to graduate from one year to the next, or to obtain a certification or license needed to enter a profession, were suddenly denied the opportunity to leverage their educational efforts to enter their profession of choice. As the crisis further developed and institutions and governmental authorities extended stay-at-home orders in 2-week increments, additional waves of cancelations were propagated. In some cases, individual examinees would no sooner find a new appointment than have that appointment canceled, only to start the process again. In the worst cases, this cycle could repeat numerous times. As we think about our own challenges managing personal, professional, and (inter)national upheaval, we have great empathy for examinees also dealing with these challenges. This was an untenable situation for examinees, testing organizations, test delivery vendors, and score users. Licensing and certification organizations and their vendors applied substantive resources to customer operations management and communications with multiple stakeholder groups and, on occasion, considerable resources into creating alternative delivery paths when plausible.

Institutions and regulatory bodies who use educational and assessment results as part of their graduation or regulation operations were forced to reckon with a backlog of students or applicants who, through no fault of their own, were unable to complete requirements either for
that have increased alongside broad antitesting and an-
other movements from student and professional groups.
This is not a new phenomenon; many professional licens-
sessing add to educational qualifications. Of course,
backlash. Organizations have faced increased stakeholder
has turned into extraordinary examinee and stakeholder
and potential pandemic resurgence, limited examinee access
routine testing continue with uncertainty around resumption
of use across the industry, in both educational and certifica-
tion and licensure examinations.
All examination programs face similar decisions and
options. However, it is important to note that the most
challenged implementations have been those that have
simply applied an existing test design to a new delivery
immediate pressure to respond. The response
must balance two extremes and their related risks. The first
extreme is not being responsive enough to examinee and
stakeholder needs under pandemic conditions, leading to di-
minished confidence in the testing organization and pres-
sure against the continued use of the services it provides.
The second extreme is being hyper-responsive, with rushed
decisions leading to potential technical failures or security
breaches and the concomitant lack of confidence in the re-
sulting scores. The most obvious first decision is to do noth-
ing: to wait for pandemic concerns to abate and for vendors to
resolve the issue by sequentially reopening previously exist-
ing testing capacity. The other choice is to create alternative
testing channels.
With test center capacity experiencing previously unseen
dramatic contraction, alternate delivery channels have
been both considered and utilized. Decision making on such
channels rely on some combination of financial considera-
tions, security concerns, technology and driver flexibility, and
exclusivity clauses in vendor contracts. Our organization has
rapidly worked with schools to establish temporary testing
sites across North America and is currently planning large-
scale event tests to simultaneously administer examinations
to thousands of examinees across the country. We are explor-
ing the use of remote proctoring, which has seen an explosion
of use across the industry, in both educational and certifica-
tion and licensure examinations.
All examination programs face similar decisions and
options. However, it is important to note that the most
challenged implementations have been those that have
simply applied an existing test design to a new delivery
immediate need to innovate.
Given current examinee backlogs and backlash, testing or-
ganizations face immediate pressure to respond. The response
must balance two extremes and their related risks. The first
extreme is not being responsive enough to examinee and
stakeholder needs under pandemic conditions, leading to di-
minished confidence in the testing organization and pres-
sure against the continued use of the services it provides.
The second extreme is being hyper-responsive, with rushed
decisions leading to potential technical failures or security
breaches and the concomitant lack of confidence in the re-
sulting scores. The most obvious first decision is to do noth-
ing: to wait for pandemic concerns to abate and for vendors to
resolve the issue by sequentially reopening previously exist-
ing testing capacity. The other choice is to create alternative
testing channels.
With test center capacity experiencing previously unseen
dramatic contraction, alternate delivery channels have
been both considered and utilized. Decision making on such
channels rely on some combination of financial considera-
tions, security concerns, technology and driver flexibility, and
exclusivity clauses in vendor contracts. Our organization has
rapidly worked with schools to establish temporary testing
sites across North America and is currently planning large-
scale event tests to simultaneously administer examinations
to thousands of examinees across the country. We are explor-
ing the use of remote proctoring, which has seen an explosion
of use across the industry, in both educational and certifica-
tion and licensure examinations.
All examination programs face similar decisions and
options. However, it is important to note that the most
challenged implementations have been those that have
simply applied an existing test design to a new delivery
approach. These copy-and-paste designs have achieved, at best, uneven success. Regardless of approach, examination programs need to engage in flexible thinking about content exposure, long-term content sustainability, uneven security concerns and mitigation strategies, and the integration of numerous dependent end-to-end technology systems and processes. The same test design will not mitigate the disparate risks of the various possible delivery channels, nor will the technology platforms across channels necessarily support the delivery of the same content. Yet, programs wishing to do more than wait for brick and mortar vendors to address the issue need to proactively tackle the challenge of delivering examinations through alternate channels.

**Future Need for Flexibility**

Once the current pandemic subsides adequately to allow for the lifting of social distancing restrictions, testing organizations will need to reflect on their pandemic actions and which, if any, of those decisions will be made permanent. For example, organizations that have relied on vendors to manage proctoring and delivery services and have attempted to stand-up alternatives that faced challenges may well gain an appreciation for the enormous challenges the logistical component of executing to create a seamless examinee experience end to end entails. Test development and psychometrics are a necessary but insufficient basis for effective assessment.

Our opinion is that there is no one correct approach to responding to the immediate challenges of the epidemic. Similarly, there is not a correct or incorrect limit on flexibility, nor a specific innovation that will resolve test delivery issues in a widespread manner. We have seen an incredible range of reactions within credentialing and licensure alone before even considering the broad range of assessments represented across NCME membership. Instead, we encourage assessment professionals to be guided by the fundamental tenets of validity. Given the inferences an assessment is designed to support, what evidence can be brought to bear to support these inferences; what threats to these inferences exist and (how) can they be effectively mitigated? Remote proctoring has become one popular solution. This may be a perfectly reasonable solution in select cases, and is almost certainly not appropriate in others, depending on factors including specifics of implementation, context, and inference to name but three. There is no magic bullet. Instead, there is an opportunity for creative interdisciplinary science to solve these challenging problems. As is the case in creating test designs now, maintaining the right line of sight in decision making will be critical: scores that are too often heavily influenced by preknowledge or other nonconstruct relevant factors will not effectively support inferences, meet customer need, nor be self-sustaining in the long run.

Regardless of actions taken during the pandemic, organizations will be remiss if they do not take the opportunity to flesh out their contingency plans for future testing disruptions. Decisions will focus around delivery channels, balancing psychometric innovation with realistic technology implementation requirements, and reexaming the trend toward continuous testing. Programs will likely cement the need for contingency plans to proactively include the ability to quickly flip the switch on multiple examination delivery paths under various emergency scenarios. These will likely include site expansions, event tests, remote-proctored administrations, and perhaps even a return of computerized examinations to temporary paper and pencil administrations. And, with the realization of the need to quickly deliver via multiple delivery pathways, validity evidence surrounding innovative technology-enhanced item types will need to be weighed against the cost and complexity to support those items across multiple technology platforms that would be leveraged under emergency scenarios. Additionally, the rise of “customer choice” for delivery channels (i.e., the ability of an examinee to choose to test via brick and mortar, event-based, or on-line remote-proctored administrations) may be accompanied by a fall in continuous testing opportunities to ameliorate security concerns that threaten score validity.

The COVID-19 pandemic has impacted all areas of our lives: where and how we work, how we socialize, and even where and how we eat. The testing industry is not exempt from disruption. The pipeline of examination stakeholders has been impacted, arguably most severely in the health professions where the pandemic has both increased the need for practitioners while breaching the pathway to practice. Testing organizations have varied in their response, yet all must be ready to respond to stakeholder inquires on their thinking around whether, and how, they plan to build additional testing capacity under social distancing guidelines and test center closures while maintaining existing validity arguments they have worked so hard to compose. The decisions made by testing organizations today will not only dictate their continued relevance to stakeholders and therefore their own survival, yet also dictate future examinee expectations, which will drive the ways in which examinations will be delivered moving forward.

**Reference**

Feinberg, R. A. (2020). Practical approaches for estimating classification decision on interrupted high stakes tests. Paper to be presented at the Annual Meeting of the Psychometric Society (IMPS); New York, NY.