October 24, 2022

VIA E-MAIL: Rulecomments@dcwp.nyc.gov

City of New York
Department of Consumer and Worker Protection
42 Broadway
Manhattan, New York 10004

Re: Letter of Comment on the DCWP Clarification for NYC Local Law 144

Dear Sir/Madam:

The Institute for Workplace Equality ("IWE" or "The Institute") submits the following comments in response to the New York City ("NYC" or the "City") Department of Consumer and Worker Protection's ("DCWP" or the "Department") invitation. The Department's Notice of Proposed Rules is seeking to clarify the requirements set forth by NYC's Local Law 144 that will regulate the use of automated employment decision tools ("AEDT") wherein hiring decisions are made or substantially assisted by algorithmically-driven mechanisms.

Background on The Institute for Workplace Equality

The Institute is a national non-profit employer association based in Washington, D.C. The Institute's mission includes the education of federal contractors as to their affirmative action, diversity, and equal employment opportunity responsibilities. Members of The Institute are senior corporate leaders in EEO compliance, compensation, legal and staffing functions representing many of the nation's largest and most sophisticated federal contractors.

The Institute recognizes the responsibility of all employers to create a nondiscriminatory workplace. To that end, NYC's DCWP has an important role in enforcement efforts related to Local Law 144 and additional rules to clarify the requirements for compliance. This is critical to ensuring that employers understand their requirements and can effectively comply with the law beginning January 1, 2023.

I.   Overview of Proposed Rules for Local Law 144

The Institute applauds the DCWP's efforts to clarify Local Law 144. Enhanced clarity established by clearly defining terms and requirements allow employers to ensure they are

compliant with this law when it goes into effect. Our members are encouraged by various clarifications in the proposed rules so far.

However, there are some areas where the proposed rules fall short of providing clarifying guidance and/or deviate from standards and accepted practices related to evaluating disparate impact. Additionally, there are areas in which inconsistencies raise questions that require further clarification to allow employers and auditors to understand how to properly comply with Local Law 144. The remainder of these comments will focus on these areas in which we encourage the DCWP to consider further revising and clarifying the provided definitions and requirements.

II.   Who is eligible to conduct a bias audit?

In §5-300, the DCWP defines an independent auditor as: "… a person or group that is not involved in using or developing an AEDT that is responsible for conducting a bias audit of such AEDT." This definition suggests that neither an employer using an AEDT nor a vendor developing an AEDT will suffice as an independent auditor. However, in the first example of § 5-301, the statement is made that:

> The employer asks the vendor for a bias audit. The vendor uses
> historical data it has collected from employers on applicants
> selected for each category to conduct a bias audit…

Understanding that this example is describing a situation in which the employer is considering the use of an AEDT (as opposed to an employer having already used an AEDT) such that no employer-specific data are available and historical vendor data may be required, the phrasing suggests that a vendor is conducting the bias audit of their own AEDT for an employer.

It is important to distinguish between who is conducting the bias audit and who is providing the data upon which the bias audit is based. There will most certainly be use cases, such as the one outlined in the DCWP's provided example, wherein an AEDT has not yet been deployed for an employer, and a vendor may need to provide relevant data to an independent auditor to be used in conducting the bias audit analyses. However, a vendor providing data for a bias audit and a vendor being eligible to serve as the independent auditor of their own AEDT and conducting the bias audit are quite different. As such, we request the DCWP clarify this issue in the proposed rules to reflect their intent. If the definition of independent auditor precludes vendors from performing bias audits, then the example should clarify that the vendor is providing data for the bias audit to an independent auditor, but not performing audits of their own tools.

III.   What data should a bias audit be based on for different use cases?

The proposed rules do not make it clear whether employers are required to use their specific data or whether cross-client, aggregate data from a vendor might suffice. We recognize that this answer may depend on the specific circumstances. In some situations, an employer will

have their specific data available to use (*e.g.*, post deployment of an AEDT, pilot studies with an AEDT), while other situations may occur where no employer-specific data are available (*e.g.*, prior to deployment of an AEDT). In situations where employer-specific data are not available, there will be some circumstances in which historical, archival data from other sources may be available, and some in which no historical, archival data are available (*e.g.*, custom-built AEDT pre-deployment) to estimate the potential bias of an AEDT.

We request the DCWP provide clarification for three typical use cases regarding the data that should be used for the purpose of conducting bias audit analyses. Specifically, we recommend that the DCWP address the following use cases and provide guidance in the proposed rules.

1. <u>Implementing an AEDT where sufficient employer data are available</u>: This scenario aligns well to the Uniform Guidelines and should most closely mirror the reality of operational use of the AEDT. When should employer data be analyzed for the bias audit?

2. <u>Implementing an AEDT where sufficient employer data are not available, but sufficient cross-employer vendor data are available</u>: This is similar to a long-standing practice by vendors to use aggregated data to estimate how an assessment would be expected to function, absent employer-specific data. When should cross-client, aggregate vendor data be analyzed for the bias audit?

3. <u>Implementing an AEDT where no existing data are available</u>: How should employers proceed when seeking to deploy a custom-built AEDT, for which no prior data exists? Are there options available for auditors to evaluate bias via a proxy method (*e.g.*, another similar type of AEDT, research literature related to a type of AEDT), or will the City grant a grace period for this particular use case?

Providing clarification on data eligible for analysis in these situations would be beneficial to employers and auditors to ensure that the bias analyses conducted are compliant with Local Law 144.

Additionally, DCWP should seek feedback as to whether the specific language used in the examples in the proposed rules are consistent with the remainder of the information contained in the proposed rules and with the intention of the City.

IV. Are there parameters that define data to be used in bias audit analyses?

After determining the source of data that are eligible for use in different situations, it is important to consider relevant parameters to conducting the bias audit. To build consistency in what employers are expecting, what auditors are delivering, and what the City is evaluating for compliance, the City would be well served by directly addressing the following parameters:

1. Candidate location: Should bias audit analyses restrict the candidate pool to include only NYC residents, or may a larger candidate pool be considered (*e.g.*, state, region, national)?

2. Relevance of data when employer-specific data cannot be used: In some situations, AEDTs may be used across a very wide range of jobs (*e.g.*, hourly workers, sales professionals, managers, executives) or in very different settings (*e.g.*, finance, manufacturing, customer service). Should bias audit analyses be restricted those jobs and settings that more closely represent the jobs for which the employer intends to use the AEDT?

V. The definition of impact ratio

We applaud the DCWP for providing more clarity around the bias audit analyses that are expected. The impact ratio is defined in § 5-300 as:

> [E]ither (1) the selection rate for a category divided by the selection rate of the most selected category or (2) the average score of all individuals in a category divided by the average score of individuals in the highest scoring category.

$$\text{Impact Ratio} = \frac{\text{Selection rate for a category}}{\text{Selection rate of the most selected category}}$$

OR

$$\text{Impact Ratio} = \frac{\text{Average score of individuals in a category}}{\text{Average score of individuals in the highest scoring category}}$$

The first equation is one of the well-recognized, practical significance metrics for evaluating selection rate differences between two different demographic groups. This metric is consistent with the "four-fifths rule" as described in the Uniform Guidelines on Employee Selection Procedures[1] ("Uniform Guidelines" Section 4D) and their Interpretation and Clarification (Questions and Answers) document[2] ("Q&A" Question 12).

The second equation, however, is not a metric that is used in practice based on our experience. It contains a fundamental flaw that does not allow for meaningful conclusions about the difference between scores on an AEDT for two different demographic groups because it does not account for variability in scores. This metric, as defined, may lead to purposefully or

---

[1] *Uniform guidelines on employee selection procedures*, Equal Employment Opportunity Commission, Office of Personnel Management, Department of Justice, Department of Labor, and Department of Treasury, 43 *Fed. Reg*. 38290-38315 (August 25, 1978). *See* 29 CFR Part 1607 (1978) available at https://www.ecfr.gov/current/title-29/subtitle-B/chapter-XIV/part-1607 and 42 CFR Part 60-3 (1978) available at https://www.ecfr.gov/current/title-41/subtitle-B/chapter-60/part-60-3?toc=1.

[2] *Adoption of Questions and Answers to Clarify and Provide a Common Interpretation of the Uniform Guidelines on Employee Selection Procedures*; Equal Employment Opportunity Commission, Office of Personnel Management, Department of Justice, Department of Labor, and Department of Treasury (1979). 44 *Fed.Reg*. 11996-12009 (March 2, 1979); available at https://www.eeoc.gov/laws/guidance/questions-and-answers-clarify-and-provide-common-interpretation-uniform-guidelines.

inadvertently illusory, misleading results. In simple terms, if a child comes home from school and tells you that they scored an 85 on a test at school, your first question might be, "Is that good?" From the information provided, you cannot tell—you lack relevant context. To know if this is a good score on the test, you need to know how others did perform (or typically perform) on the test—you need to understand the distribution (variability) of scores on the test.

That said, we very much appreciate that the DCWP has recognized the need for a bias indicator to handle the situation wherein an AEDT is not being used with a strict threshold (*e.g.*, cut-score, band, categorization) that allows for computing a selection rate, but instead scores are generated and are being directly used as a factor in determining which candidates proceed in a hiring process. This is a very common use case, and we agree that it would be best if a consistent method were used across auditors.

To that end, we submit that there are multiple, professionally-accepted, practical significance measures that may be used for effectively evaluating the difference between two "average scores" on an AEDT, and that the DCWP consider adopting one of these. For example, one such statistic that is commonly used is the d-statistic.[3][4] In simple terms, this metric allows one to describe the difference in magnitude of two average scores, regardless of the distribution of scores, in a consistent, meaningful, and interpretable way.

To compare this to the second impact ratio metric described in the proposed rules, consider an example of three different AEDTs where the average score of two groups is being compared. Then consider how the interpretation of results is much different once you have information about the distribution of scores (*i.e.*, the variance or standard deviation). The second impact ratio defined in the proposed regulations is represented by the "Averages Ratio" column, the variability of scores in the distribution is represented by the "Standard Deviation" column, and the result of calculating the standardized difference in average scores is represented by the "d-statistic" column.

Scenario	Average Group 1	Average Group 2	Averages Ratio	Standard Deviation[5]	d-statistic
AEDT 1	50	44	.88	24.0	0.25
AEDT 2	50	44	.88	12.0	0.50
AEDT 3	50	44	.88	6.0	1.00

In all three cases, the Averages Ratio appears to be the same—one could draw the same conclusion about the difference in demographic group scores. However, when considering the distribution of scores with a statistic such as the d-statistic, it is clear that the same conclusion should not be drawn. With AEDT 1, the difference is relatively small because there is a large

---

[3] Cohen J. (1988). Statistical Power Analysis for the Behavioral Sciences. New York, NY: Routledge Academic.
[4] The d-statistic is sometimes referred to as an effect size difference or a standardized mean difference.
[5] This represents a pooled standard deviation which is commonly used for this statistic. This provides a more accurate representation of the overall standard deviation for the pairwise comparison of group 1 and group 2 averages.

amount of variation in scores; with AEDT3, the difference is large as there is little variation in scores. The interpretation of bias for these two AEDTs would be quite different.

It should be clear from this example that the currently-defined second impact ratio is not very meaningful and will be unlikely to yield interpretable results. Given this, we request that the DCWP define a practical significance metric commonly used or professionally-accepted that will be more appropriate for use in comparing the average score of two groups, of which the d-statistic is but one example.

VI. Parameters to clarify related to calculations of an impact ratio

After defining the impact ratio, there are practical parameters to a bias audit analysis that are important to consider. These parameters are related to the stability of results and the confidence one might have in drawing conclusions based on the analyses conducted. We recommend the DCWP consider addressing some key parameters in the proposed rules or indicate that there is allowance for reasonable professional judgment by the independent auditor.

Note that some of these parameters may be informed by the Uniform Guidelines and their Q&A document since the proposed rules for Local Law 144 state that:

> These calculations are consistent with Section 1607.4 of the EEOC Uniform Guidelines on Employee Selection Procedures. *See* 29 CFR §1607.4.

1. Sample size required for a pairwise comparison: Small sample sizes are especially problematic for impact ratio analyses. Evidence based on small numbers is considered in the Uniform Guidelines (4D) and their Q&A document (e.g., Question 20, Question 21). For example, in Question 20 they provide an example whereby an impact ratio based on a pool of 20 males and 10 females, where 3 males and 1 female are selected, yielded a conclusion that "the number of selections is too small to warrant a determination of adverse impact." In Question 21, they provide an example in which impact ratio results would change appreciably with the shift of a single individual, noting that this indicated results that were unstable. Given the nature of impact ratios, this instability at small sample sizes is to be expected—the influence of a decision made about a single individual may unduly effect the statistical conclusions.

   We request the DCWP directly acknowledge the relevance of the federal guidance, directly specify guidance on minimum sample sizes to warrant analyses, or state that this is subject to independent auditor discretion.

2. Sample size required for a demographic group to be evaluated: We would appreciate the DCWP confirming that the guidance set forth by the Uniform Guidelines Section 15A-2[6] and Question 16 of their Q&A document, indicating that groups which

---

[6] 29 CFR §1615(a)(2); 42 CFR § 60-3.15(A)(2).

comprise less than 2% of the pool should be excluded from analysis, should be applied when conducting Local Law 144 bias audits.

3. <u>Interpretive guidelines for impact ratios</u>: Interpretation of an impact ratio has become more nuanced over the years. Impact ratio results can be interpreted in a binary framework based on the four-fifths rule, or with a multi-categorical interpretation (*e.g.*, OFCCP Q&A on Practical Significance, January 15, 2021[7]), or with more of a continuous consideration taking the statistic as a way to gauge the potential for a tool to yield bias. Will the DCWP provide interpretation guidance, or will this be left to the discretion of the independent auditors?

4. <u>Which impact ratio to calculate</u>: §5-301-b indicates that the averages ratio must be calculated when candidates are scored. However, in the vast majority of cases, any candidate that has been assigned to a classification or band has been given a score upon which the assignment is based. As such, the proposed rules read that a classification or banding approach will require both be calculated.

   Given the definition of impact ratio in §5-300, which clearly indicates (via the "OR") that the intent is to calculate one metric (not both), it would behoove all for the DCWP to align language in the proposed rules to reflect that only one metric should be calculated and that this should be the one that most closely mirrors the operational implementation of the AEDT.

VII. Which pairwise comparisons are required for analysis?

The tables in the two examples in the proposed rules imply intersectional analyses need to be conducted and reported (*i.e.*, the combination of race/ethnicity and sex). While intersectional analyses can be conducted, in most cases the primary interest is in evaluating race/ethnicity and sex separately, with intersectional analyses considered as supplemental at most.

This is another situation in which it would behoove all involved for some clarification on the pairwise comparisons that are intended to be evaluated in a bias audit for Local Law 144. Should auditors analyze race/ethnicity and sex separately only, analyze intersectional combinations only, or do both (yielding three tables per analysis strata: 1 race/ethnicity, 1 sex, 1 intersectional).

VIII. Effective date for Local Law 144

NYC's Local Law 144 is slated to go live on January 1, 2023. However, there is still significant confusion related to the specifics for employers coming into compliance with this law. Further, as noted here, there are some additional clarifications that are necessary and will impact the ability for employers and auditors to comply with what is required. Finally, with the

---

[7] https://www.dol.gov/agencies/ofccp/faqs/practical-significance

rescheduling of the October 24 public hearing to a later date, both the feedback to the DCWP, as well as any revisions to the proposed rules based on this feedback, will be further delayed. We respectfully request that the City immediately consider a six month delay to the effective date. In the alternative, the City could proceed with a soft launch in which there is a grace period for employers to comply, so long as an independent auditor is identified and the work is scheduled by January 1, with a requirement to have the work completed before the year concludes.

Conclusions

Thank you in advance for your consideration of The Institute's comments. We are happy to provide any additional information you may need or to answer any questions you may have.

Best wishes,
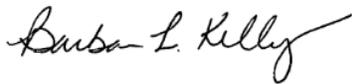
Barbara L. Kelly
The Institute for Workplace Equality Director