Emily Thaden, Ph.D.
Director of National Policy & Sector Strategy
PO Box 42255
Portland, OR 97242
503-493-1000 ext 3

June 5, 2017

Mr. Mel Watt, Esq.
Director
Federal Housing Finance Agency
400 7th St SW
Washington, DC 20024

RE: Comments on the Duty to Serve Evaluation Guidance 2018-2020 Plan Cycle

Dear Director Watt & your esteemed colleagues,

I write to you today to express my comments and concerns on the proposed Evaluation Guidance (hereinafter "Guidance") that will be used to evaluate Fannie Mae's and Freddie Mac's Underserved Market Plans to fulfill obligations under the duty-to-serve rule. I would first like to thank FHFA for completing the final rule and supporting a meaningful public comment process to inform rule-making and Plan development.

I hope that FHFA will consider input on the Guidance as deeply as it has on the development of the final rule. In light of the fact that the proposed Underserved Market Plans (hereinafter "Plans") have been designed based upon the proposed Guidance—and public comments on the Plans are due before the Guidance is finalized—I fear that public input may have less of an impact on the Guidance than it should. However, I strongly believe that the Guidance, as proposed, requires substantial changes.

I hold an interdisciplinary doctorate in the social sciences, and I have extensive training and education on survey and scale design as well as statistics. My expertise in these areas informs my concerns over the proposed evaluation process. I will delineate below the issues I have identified in the proposed evaluation process and some suggestions for improvement. Please know that I would be eager to provide constructive suggestions and ideas for how the evaluation process may be improved during revisions to the Guidance. I am confident that the Evaluation Guidance can be substantially improved to: 1) have less of an adverse influence on the Enterprises' Underserved Market Plans, and 2) ensure that the evaluation process is both more comprehensible and translatable. This can be done while also complying with the final rule.

**Concerns & Recommendations**
In general, I am concerned that the Guidance is often confusing, contains vague measures, and disincentivizes more impactful objectives by the GSEs. My specific concerns and suggestions follow.

1. **Passing or failing duty-to-serve obligations should account for impact, not solely the extent to which objectives are accomplished.** Simply put, an Enterprise should in no way be able to pass their duty-to-serve obligation by accomplishing trivial objectives; however, this is absolutely possible under the proposed Guidance. The proposed Guidance only considers the extent to which objectives are accomplished in order to determine whether an Enterprise will pass or fail their duty-to-serve obligation for a specified Underserved Market. Logically, this should influence the Enterprises to submit easy-to-do, low risk, less innovative, and more internally-controlled objectives. Why? Because passing or failing currently depends simply upon accomplishing objectives, so one is not incentivized to propose objectives that are hard to accomplish. For that reason, Step 1 should include an assessment of the impact of objectives in order to determine whether the Enterprises failed or succeeded at fulfilling duty-to-serve obligations. It is inadequate to leave the assessment of impact— or put differently, how substantial and meaningful the objectives are to transforming an Underserved Market—only in the relative rating of how well an Enterprise passed per Step 2.

   I recommend the incorporation of an impact measurement into any assessment of pass/fail that is evaluated in a separate and discrete way from the extent to which objectives are accomplished (my comments and recommendations on that are below). I also provide more detail on the problems with the current impact assessment and recommendations for a better evaluation process below.

2. **The Step 1 scale incentivizes easy-to-accomplish objectives.** Completely separate from my comments in Number 1 above, the use of an ordinal scale, rather than an interval scale, lacks logic. Despite the proposed Guidance stating the opposite intent, the scale and calculation of a Total Score for Step 1 discourages the Enterprises from being ambitious in their objectives. To explicate, only objectives that receive a 10 (i.e. fully accomplished or exceeds target) pull up the average for the Total Score. Therefore, any objective that is challenging or looking as though it may not be fully accomplished in the plan cycle, should arguably be abandoned and effort should be redirected towards objectives that will receive a 10. In particular, logically the Enterprises should abandon objectives if they have completed 25% of the objective but not 75% of the objective since this would require 50% of the work on the objective for only a 3 point increase in the objective's score, which ultimately will still pull the average down for the Total Score of 7 that is needed to pass.

   Additionally, it is poor practice in scale design to imply meaning in numbers that is not actually there. For instance, it is misleading that the 0-10 scale does not actually utilize an 11-point rating scale nor that 6 equates to 60% or that 3 equates to "moderate."

   It is understandable that there are restricted capabilities to measure the extent of accomplishment on a more refined scale, especially for outreach and loan product objectives. However, scale construction for Step 1 can easily be improved by applying a 5-point interval rating scale:

   | | |
   |---|---|
   | 4 | 76 through100% Accomplished |
   | 3 | 51 through 75% Accomplished |
   | 2 | 26 through 50% Accomplished |
   | 1 | 1 through 25% Accomplished |

| | |
|---|---|
| 0 | 0% Accomplished |

The average of all scores on objectives in each Underserved Market would need to equal or be greater than 2.875 (or 71.875%) in order to pass Step 1. This scale minimizes the level of "punishment" or disincentive for incorporating more challenging objectives or those with less internal control that may or may not be fully accomplished.

3. **Step 2 conflates too many factors for accurate assessment.** Simply stated, Step 2 is attempting to account for too many factors into one score presented in Appendix B. Too many dimensions and factors are being combined for an accurate assessment. If one analyzes Appendix B, all of the following dimensions and factors are present: direct/indirect impact, current/future impact, extent of impact, efficiency in implementation, effectiveness of implementation, difficulty to accomplish, and extent accomplished. An example of one problem this poses is duplication in the "Implementation" factor in Step 2 and the "Extent to Which Objective was Accomplished" factor, which is measured in Step 1. For instance, under a score of 10 in Appendix B of Implementation, the description reads, "Enterprise implemented at least nominally the actions associated with the activity," which clearly speaks to extent accomplished (not how efficiently the Enterprises were at whatever they did accomplish). Meanwhile Implementation and Impact are conflated under a score of 0 for Implementation: "However implemented, the objective represents less than a minimal impact…" Both of these descriptions duplicate measures of impact or extent achieved, and they do not measure what the Guidance says it should, which is how effectively or efficiently the Enterprises implemented the objective. In terms of the Implementation factor, the descriptions in Appendix B are clearly accounting for impact as well as difficulty, which is problematic. These factors should not be conflated.

   Ultimately, due to the factors *not* being mutually exclusive—which is best practice for objectively scoring qualitative data— it is incredibly unclear how to consistently determine a Score. For instance, if an objective is only 50% accomplished but it is high impact, highly difficult, but efficiently implemented (relative to the level of difficulty), what would the Score be? If an objective was 75% accomplished, but it was very low impact and not efficiently accomplished, what would the score be? What happens if the Enterprises proposes activities to accomplish an objective, but they accomplish it in a different way? How would this affect the scoring?

   Other issues with Appendix B include that the Guidance states that an objective must meet both impact and implementation criteria for a specific score in order to receive that score. This fails to recognize that there may be an acceptable inverse relationship between impact and efficiency. In other words, something may have substantial impact but not be particularly efficient to implement or vice versa. Next, it appears that it would be very hard to score highly on objectives that were not fully accomplished per the descriptions in Appendix B, which is inherently problematic since extent of accomplishment should be fully determined by Step 1 and not unfairly continue to penalize the Enterprises for activities that were harder or less certain of being fully accomplishable. This dissuades harder-to-do, ambitious objective-setting in the Plans.

   Lastly, this scale is also problematic and does not abide by best practices for scale design. When applying qualitative ratings, it is vital that each rating is clearly articulated and defined in an objective and mutually exclusive way from: 1) other factors, and 2) other ratings. As reviewed above, the factors are conflated and the ratings are not clear either. Terms like "minimal" (score 10), "meaningful" (score 30), and "comprehensive" (score 50) are subjective; therefore, defining what constitutes a score of 20 or 40 is nearly impossible, and I do not believe there would be positive inter-rater reliability (i.e. strong agreement across different raters)—a key indication that a scale is working well. Once again, there is no point to having a scale from 0-50 when there are only 6 possible ratings.

I recommend that Impact is scored as a completely independent factor from extent accomplished. Then, the score for each objective can get a boost for level of difficulty or extra credit, and conversely, can get a penalty for inefficient execution (because ultimately, isn't the primary point of Implementation in Appendix B to assess whether the Enterprises poorly used resources or ineffectively pursued the objective?). This would ensure that various factors are teased apart and independently assessed without duplication. Notably, I believe that the impact score would work better not as an average but as a sum so that accomplishing many lower impact activities is not penalized more than doing fewer high impact activities.

4. **The Step 2 averaging procedure results in biasing the proposed Concept Scores for objectives.** An objective assigned a score of 0-20 will be lower priority to accomplish because one can't score higher for over performing. Therefore, small but meaningful tasks will be disincentivized. Unsurprisingly, what we see in the proposed Underserved Market Plans is that hardly any (if any) activities are proposed to have a Concept Score below 30, rendering the very intention of a Concept Score scale of 0-50 effectively useless. The reason for this is because the Enterprises would be dissuaded from being forthright about activities that are important but not necessarily large impact (or difficult), because these would actually injure their overall performance and final score. Hence objectives, such as outreaching to handful of lenders and purchasing a lot of loans, are given the same proposed Concept Scores.

   Furthermore, FHFA needs to be clearer in Appendix B about scoring for completed loan purchases. Under the score of 30, it states "loan purchases will be assumed to be effectively implemented", and under the score of 50, it states that "all loan purchases will be assumed to qualify for a score of 50 under the implementation criterion." What does this mean and how does this interact with the Guidance for assessing impact? Clearly loans can be purchased in a poorly implemented fashion and depending upon the objective set and extent of accomplishment, the impact of loan purchases may be poor. Therefore, I am unclear what FHFA's rationale is for this. I hope this will be reviewed and explicated in the final Evaluation Guidance.

   Arguably, the Enterprises shouldn't be punished for conducting objectives that are important but not necessary extremely impactful, as these may be important to laying the foundation for higher impact objectives later. Furthermore, the Enterprises should have a choice of whether they pursue a greater number of less impactful objectives versus a fewer number of highly impactful objectives. It is the fault of the proposed Guidance to introduce unfair bias by averaging scores for objectives. This is not an appropriate use for averaging, as averaging ultimately implies that the bottom half of the scale is negative while the top half of the scale is positive. However, lower impact activities are not inherently negative, and in fact, at times necessary. Hence a summative approach to impact would make much more sense. As discussed above, assessing the effectiveness or efficiency of implementation should be done separately in my opinion.

5. **Step 2 procedure of weighting by areas is duplicative and dissuades some high-impact work.** I whole-heartedly disagree with the system of weighting by evaluation area. First of all, this introduces a step that really ruins transparency or the ability to easily translate objectives' concept scores into the effect it will have on final scoring. Second, depending upon the objective or submarket within an Underserved Market, it is not necessarily the case that loan purchase or loan product areas are always far more important (basically, doubly important) than outreach or investments.

   What matters is not necessarily the *type* of objective (i.e. evaluation area) but how meaningful and challenging the objective is. Ultimately, this is what impact and level of difficulty should be measuring. However, presently, the proposed Guidance is inappropriately assessing the importance of an objective through weighting on Evaluation Areas, which is also duplicating some of what would be reflected in a Concept Score. Once again, this process has not been designed for mutually exclusive measurement and scoring. As can be seen in the drafted Underserved Market Plans, this is clearly disincentivizing the Enterprises from making "investments and grants," and it is encouraging them to inflate the proposed Concept Scores on "outreach."

6. **Extra credit weighting is confusing.** The proposed Guidance states that even if the Enterprises include other objectives related to an activity that received lower scores, an individual objective with a score of 40 or more would qualify for an extra credit adjustment. This seems inappropriate in light of the fact that the Extra Credit is assessed on the Total Score for Step 2. Hence, why would one relatively small objective in a list of many objectives potentially be worth more to the overall Step 2 score than, let's say, an impactful investments and grants objective? As mentioned above, I believe that FHFA should revisit their relative weighting for statutory evaluation areas and extra credit. The logic behind the current proposed evaluation process is very unclear and seems to run counter to the intention in this statement, "Setting a final score of 40 as a threshold also helps FHFA ensure that extra credit eligible activities do not receive disproportionate weight in determining the final rating for the particular underserved market."

7. **Step 2 final scores are not substantiated and are seemingly very low.** Lastly, what is the rationale for the Final Rating and the various determinations (i.e. minimally passing, low satisfactory, high satisfactory, exceeds)? The Final Rating seems incredibly low. For instance, an Enterprise could receive scores across all evaluation areas of 20s and receive a "Low Satisfactory". Alternatively, an Enterprise could receive 20s for loan purchase, loan product, and investments or grants, a 40 for outreach, and 10% extra credit due to outreach and receive a "High Satisfactory." An Enterprise could receive a 30 for loan purchase, outreach, and investment or grants and a 50 for loan products and receive an "Exceeds" rating. I believe that the final scores feel incredibly low, and the system for evaluating what would be reasonable Final Rating determinations feels hard to discern due to how the current evaluation process is structured.

To conclude, I believe that this Evaluation Guidance is substantially flawed and is adversely influencing the Enterprises planning so that they propose "safe" and less impactful objectives. Furthermore, I believe the Guidance does not adequately support a transparent, objective, and fair evaluation by FHFA. I would welcome an opportunity to work with FHFA on correcting for these errors in any way I can. In fact, I have started to draft an alternative proposal that complied with the regulations, but I need more time to refine it. Thank you for being willing to hear my concerns, and I look forward to reviewing the final Evaluation Guidance.

Sincerely,

Emily Thaden