PREDICTORS OF UNREPORTED INCOME: TEST OF UNREPORTED INCOME (UI) DIF SCORES

Dennis Cyr, Internal Revenue Service
Thomas Eckhardt, Internal Revenue Service
Lou Ann Sandoval, Internal Revenue Service
Marvin Halldorson, Internal Revenue Service
Prepared for the Internal Revenue Service Research Conference, June 11-12, 2002

Nearly $60 billion per year in tax revenue is lost from the tax due on income underreported by sole proprietors and informal suppliers according to estimates of the individual gross tax gap. Unreported income is the single largest component of the gap. The difference between income that was reported voluntarily and income that should have been reported is the definition of unreported income. Both income and self-employment taxes are lost when these individuals inaccurately report their income.

Detecting unreported income is difficult. Various efforts have been undertaken by the Internal Revenue Service over the years to address unreported income. These include the Information Returns Program (IRP), the Economic Reality Ratio (ERR) and the Unreported Income Discriminant Function (UI DIF).

The Office of Examination Planning and Research originally developed the formulas for the UI DIF. Taxpayer Compliance Measurement Program (TCMP) Phase III, Cycle 6 data (Tax Year 1974 returns) were used initially. The scores were intended to identify individual income-tax returns having a high probability of unreported income. The formulas were updated using the TCMP Phase III, Cycle 9 data (Tax Year 1985).

In 1998, the Denver office of Small Business/Self-Employed Research met with the National Headquarters Office of Research and requested that the UI DIF formulas be "refreshed" with the TCMP Phase III, Cycle 10 data (Tax Year 1988 returns). In 1999, Denver and NHQ Research conducted preliminary tests of the "refreshed" formulas using data from closed examinations and developed a proposal for more extensive tests.

In 2001, Denver and NHQ Research met with the SB/SE Examination Centralized Workload Selection and Delivery group and Examination Reengineering teams to develop a plan to test the utility of the UI DIF formulas for improved workload selection. Preliminary tests were conducted on Activity Code 537 (C-TGR $100,000 and over) returns. Denver and NHQ Research then developed a prospectus for an expanded, two-phase study of the UI DIF formulas. In 2002, both phases of the study were completed.

In this study, expert classifiers were used to classify returns with high or low UI DIF scores. The purpose of this test of the UI DIF formulas was to ascertain the ability of the unreported income discriminant function (UI DIF) scores to distinguish between individual income tax returns that classifiers selected as having a high probability of unreported income and not selecting those that did not. The scores were tested in all eight of the Examination Activity Codes for which the formulas have been developed. Previously, testing had been limited to only one of the Activity Codes 537[1].

[1] Interim Report, Predictors of Unreported Income UI DIF Study, Dated August 13, 2001

The purpose of this report is to provide the customer with our final results from the two-phase study of the UI DIF formulas. We will discuss our research methods, findings and conclusions. The report will provide recommendations to our customer based on the research performed.

Research Methods

Whether or not a return was selected by trained classifiers was the criterion variable for testing the formulas. We can paraphrase this to answer the questions: Does UI DIF generally give a high score to cases that classifiers select as having the likelihood of unreported income and Does UI DIF generally give a low score to cases that classifiers accept as not having the likelihood of unreported income? Eleven experienced revenue agents were each asked to classify the same 400 returns, 50 from each of the eight Activity Codes. Each set of 50 returns had been randomly sampled from two subsets determined by use of the array of UI DIF scores: 25 returns were taken from the top two percent and 25 returns from the bottom 50 percent. The classifiers were asked to indicate whether or not they believed that any particular return exhibited the likelihood of unreported income. The 25 high and 25 low UI DIF returns from within each activity code had been randomly scrambled together. The classifiers did not know that the returns might exhibit differences. The underlying assumption in the use of this dichotomous variable as the criterion for testing the UI DIF scores is that the knowledgeable classifiers would be able to make the correct distinction between those returns that showed the likelihood of unreported income and those that did not. Against this standard, the scores could be evaluated.

The classification process for this study consisted of two phases, in an effort to provide some indication of the utility of providing classifiers with alternative case-file resources. In phase I, the classifiers did not have the original returns to observe. Instead, they only had printouts from the Midwest Automated Compliance System (MACS) for the primary tax year and two prior years. In phase II, the classifiers had, in addition to the MACS prints, the original returns for the primary year and other IRS supplied internal and external data. The inclusion of two criterion variables for the UI DIF testing necessitates two parallel analyses of the resultant data, one analysis for each phase. In both phases the same 11 classifiers were used.

Each of the classifiers looked at all 400 returns (50x8 activity codes) in both phases. The study was monitored and classifiers were instructed not to discuss their observations among each other. The instructions given to the classifiers for each phase can be found in Appendix A, Exhibits A and B. For phase I, the classifiers recorded a check in the appropriate box for "yes" or "no" for the likelihood of unreported income. In phase II, they used a classification checksheet with a box in the upper left-hand corner, which they checked if they believed there to be the likelihood of unreported income. Then, after their determination for the unreported income, the classifiers were asked to classify any other questionable or suspicious compliance item as part of a separate project for an exam reengineering team. Copies of the data capturing instruments for both phases can be found in Appendix B, Exhibits C and D.

The raw data consist of counts of assigned membership within predefined groups of interest, such as the existence or absence of a particular attribute. The data are usually presented in tables. Often the underlying variables of interest are dichotomous; that is, there are two and only two possible outcomes for the categorization, such as “yes” or “no,” “high” or “low,” etc. When there are two dichotomous variables being compared, a “2x2” table is generated. Thus, four cells are determined, such as “yes” and “high,” “no” and “low,” etc. The row and column totals are known as “marginal” totals and they sum to the “grand,” or overall total.

The structure of the analysis of the UI DIF test data can be visualized in a two-by-two table, with the two pairs of foils identifying:

(1) Whether or not a particular return was selected by a classifier as likely exhibiting unreported income and

(2) Whether the return had scored among the top two percent of returns in its activity code or among the bottom fifty percent.

Figure 1 shows a representative 2x2 table for the analysis of the data for any particular activity code that were observed during one of the two phases. The 25 high-scored returns and the 25 low-scored returns were either selected for suspicion of unreported income or were accepted as filed by the classifiers. The results were then combined as was appropriate for the analysis.

FIGURE 1
REPRESENTATIVE TABLE FOR THE DATA ANALYSIS

Phase				
Activity Code		Non-Selected	Selected	
53n	High	Y	XX	25
	Low	YY	X	25
		YYY	XXX = 50 - YYY	50

The first variable, the criterion, is presented as the columns of the tables, with “selected” to the right; and the second variable is presented as the rows, with “high UI DIF score” at the top. Consequently, an entry in the southeast quadrant or in the northwest quadrant demonstrates consistency, while the opposite corners, or quadrants, demonstrate inconsistency. Again, the columns represent the IRS’s best estimate of the accuracy of the income reporting on each return, against which the predictive quality of the UI DIF scores may be evaluated.

The cumulative counts of the selections for the collective efforts of the 11 classifiers are presented later in this document. The marginal totals for any one classifier for the predictor variable must be 25; however the marginal totals for the criterion variable may vary anywhere from zero to 50, depending on the classifier’s perceptions. The grand total within each activity code will always be 50.

Again, the research question that this study intends to answer is whether or not the UI DIF score mimics classifiers' selections regarding the likelihood of unreported income. Fisher's Exact Test and the Odds Ratio Calculation were considered to be appropriate for analyzing the UI DIF test data.

The test statistic is calculated by comparing experimental observations to expected outcomes under some underlying hypothesis. The null hypothesis for Fisher's Exact Test is that there is no relationship between the two variables, UI DIF score and select/non-select by classifier. Alternatively, the researcher anticipates that the null hypothesis will be rejected, thus confirming that there is a statistically significant consistency—that the two variables identify returns in the same way. Consistency of the northeast-southwest orientation described above is needed to support the usefulness of supplementing traditional classification with the UI DIF scores. The null hypotheses must be rejected, and the direction of rejection must be appropriate.

The values in the cells of the tables are the counts of the subjects that exhibit the characteristics or attributes of interest. For example, in this UI DIF study, one of the four cells would contain the number of income tax returns that the classifiers believed reflected the likelihood of unreported income (UI) and that had received a UI DIF score in the upper 2 percent of that particular activity code. The classifiers had said, "yes, select this return" and the UI DIF score was "high." The important thing to note is that the numbers in the cells are integers, or whole numbers, because the data consist of actual counts. Note, however, that the expected frequencies in the cells can be fractional numbers, because they are calculated.

The data that were collected in each phase came from the 11 classifiers that made "yes" or "no" distinctions for each of 50 tax returns from each of the eight activity codes[2] for which the UI DIF scores have been developed. The 50 returns in each class were evenly divided between high- and low-scored returns. The 8 classes were evaluated separately and independently. Summaries of the data were prepared. There are eight separate models in each phase; therefore, there will be 16 tables of analysis (2 phases with 8 activity codes in each.)

The data for each classifier was recoded into a one or zero for purposes of the analysis. A zero stood for "accept, no likelihood of unreported income" and a one stood for "select for the likelihood of unreported income". A level of agreement was then determined by using the majority. That is, if 6 or more of the classifiers deemed that a return should be selected, then that return was selected; otherwise, it was not. When at least six of the classifiers selected as having the likelihood of unreported income, a "1" was assigned to that case. If majority did not select, then that case was coded as a zero. Each activity code for each phase was cross-tabulated into a contingency table.

[2]Activity Codes:

532 TPI $25,000 Under $50,000
533 TPI $50,000 Under $100,000
534 TPI $100,000 and Over
535 C-TGR Under $25,000
536 C-TGR $25,000 Under $100,000
537 C-TGR $100,000 and Over
538 F-TGR Under $100,000
539 F-TGR $100,000 and Over

For both phase I and II, each activity code was then evaluated using Fisher's Exact test and the Cross-products ratio. The Fisher's exact test allowed us to either reject or retain the null hypothesis explained earlier. For the purposes of the UI DIF study, the cutoff used was .05. If Fisher's computed value was less then .05, then there was deemed to be a dependency between the UI DIF score and selection. This means that there was an association and it did not occur at random

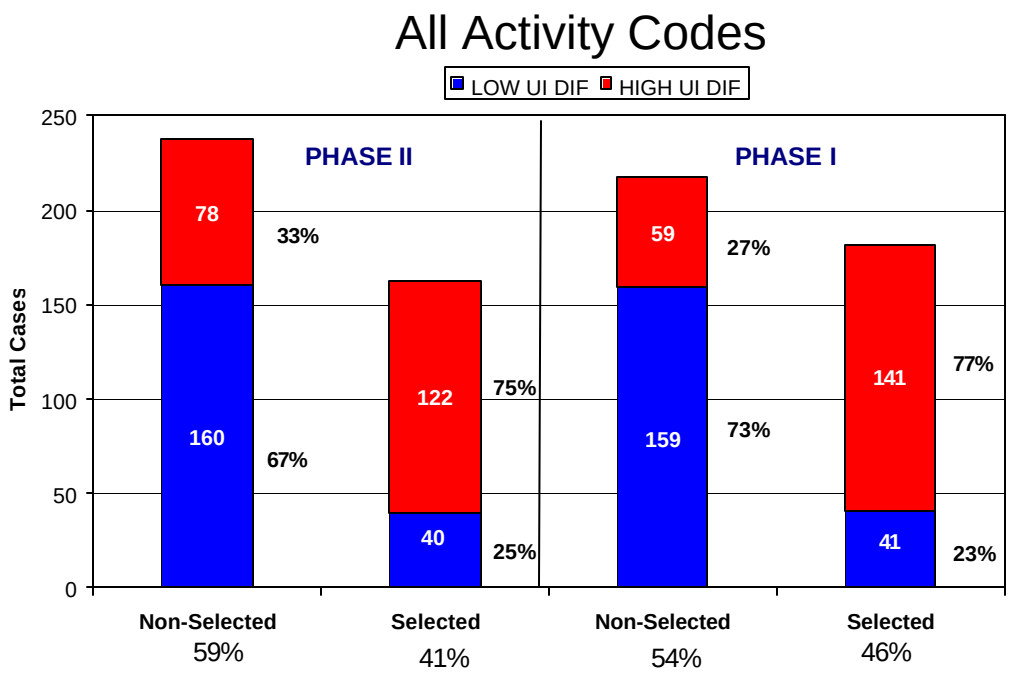
After rejection of the null hypothesis, the level of association was determined. Rejecting the null hypothesis alone, does not tell the reader if there was a weak or strong association between the score and the selection. The cross-product Ratio was computed when appropriate to show the level of association. For purposes of our testing, a "1" was substituted in the 2X2 table if any of the cells contained a zero. This is taking a more conservative approach, by placing a one in the cell for computation of the cross-product ratio. The further the ratio is from 1, the stronger the association.

The log-odds ratio was then computed to take the analysis one step further. This ratio showed the direction of the association for the study. The log-odds ratio is symmetric around zero; therefore, the researcher wanted the direction of the association going to the positive rather than the negative from zero.

Research Findings

Overall, there was little difference between the results of the two different phases of the study. The overall selection rates using majority basis for the likelihood of unreported income were quite similar, and the conclusions reached from the data are the same. Figure 2 shows the counts and percentages for each phase of selected and non-selected returns by high and low UI DIF scores combined over all of the eight activity codes for which UI DIF score formulas have been developed.

FIGURE 2
SELECTED AND NON-SELECTED RETURNS BY HIGH AND LOW UI DIF SCORES:



The results for the eight activity codes by their essentially different sources of income are somewhat unique. The following tables provide more detail from which the differences can be seen by activity code.

Activity codes 532, 533 and 534 UI DIF scores yielded inconclusive results because there were few selections made from these activity codes. Only a small proportion, ¼ or less, of the returns was selected. Of the returns selected, all selections were high UI DIF scored. No low UI DIF scored returns were selected. Figures 3, 4 and 5 show the counts and percentages of selected and non-selected returns by high and low UI DIF scores for these three activity codes, for both phases.

FIGURE 3
SELECTED AND NON-SELECTED RETURNS BY HIGH AND LOW UI DIF SCORES:

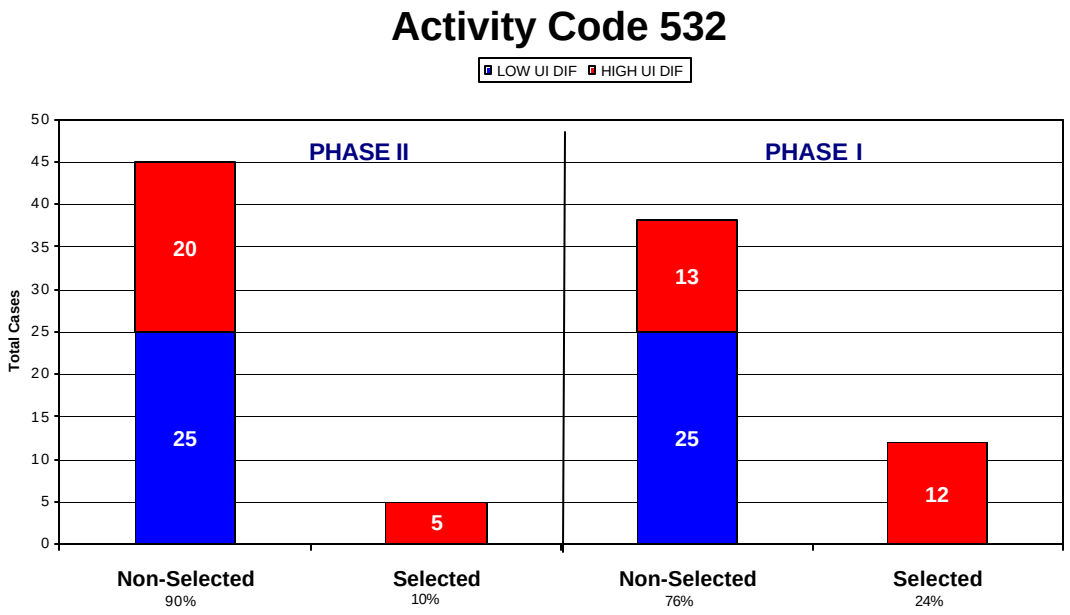


Figure 4
SELECTED AND NON-SELECTED RETURNS BY HIGH AND LOW UI DIF SCORES:

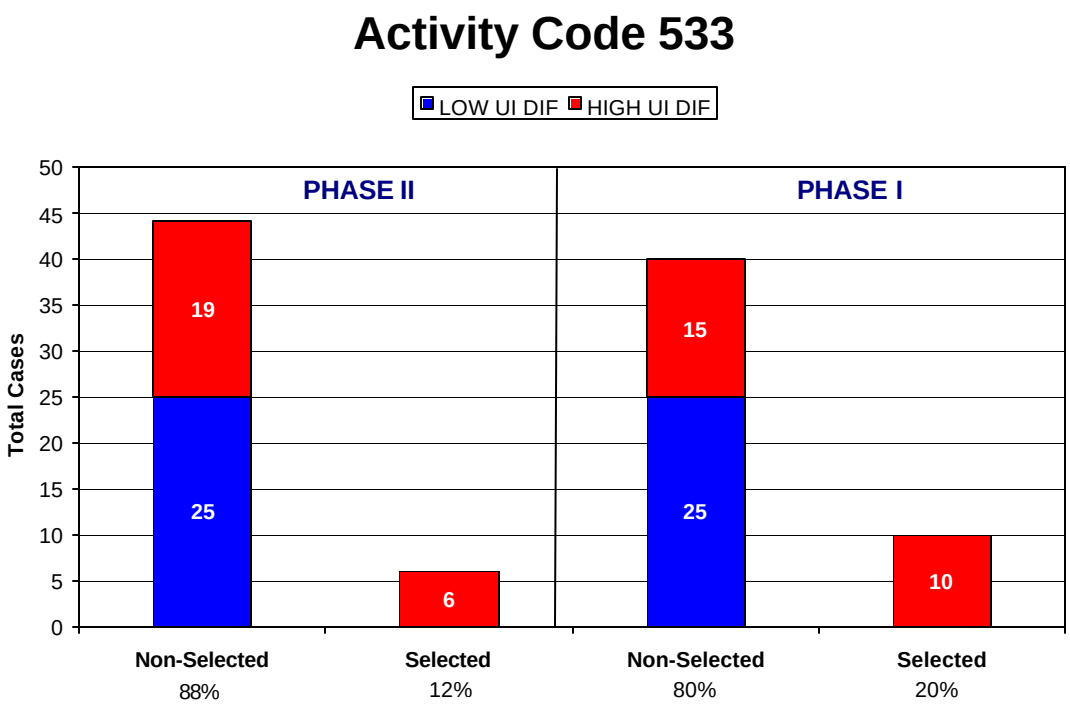
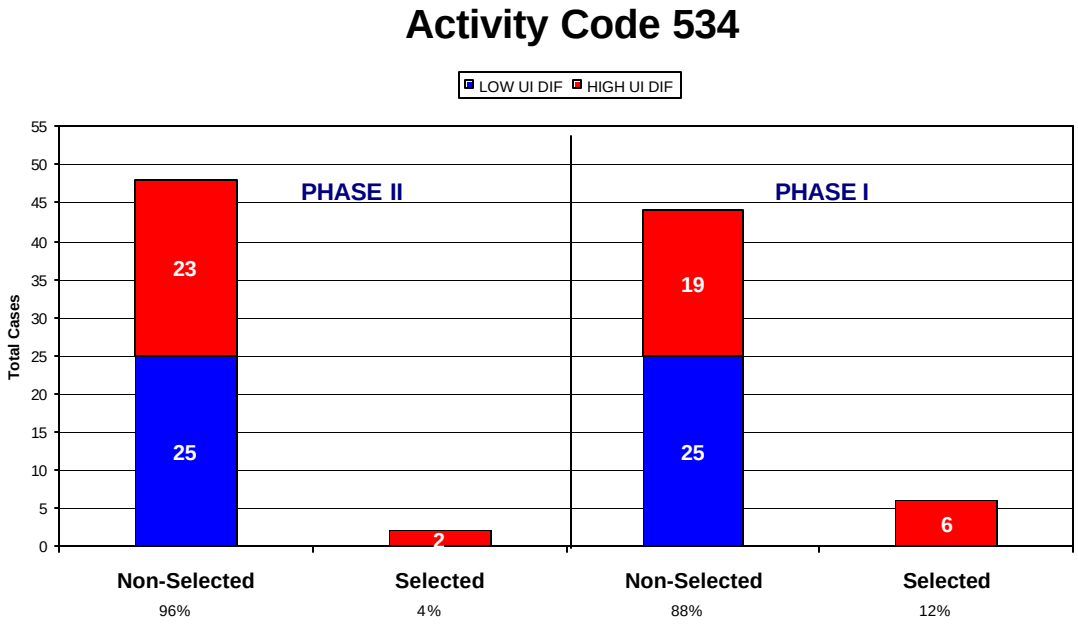


FIGURE 5
SELECTED AND NON-SELECTED RETURNS BY HIGH AND LOW UI DIF SCORES:



The results for the remaining five activity codes (535-539), were themselves somewhat similar; however, they were noticeably different from the 532, 533 & 534 activity codes. The classifiers selected higher proportions of the returns in activity codes 535, 536 and 537. In activity codes 535 and 536, 2/3 or more of the returns were selected by the classifiers as suspicious; however, although about 2/3 of the returns that were selected were high UI DIF

scored, the other 1/3 of the selected returns were low UI DIF scored. In activity code 537, just a little over ½ of the returns were selected by the classifiers as suspicious; however of the returns selected, over 80% were high UI DIF scored. Figures 6, 7 and 8 show the counts and percentages of selected and non-selected returns by high and low UI DIF scores for activity codes 535, 536 and 537 in each phase.

FIGURE 6
SELECTED AND NON-SELECTED RETURNS BY HIGH AND LOW UI DIF SCORES:

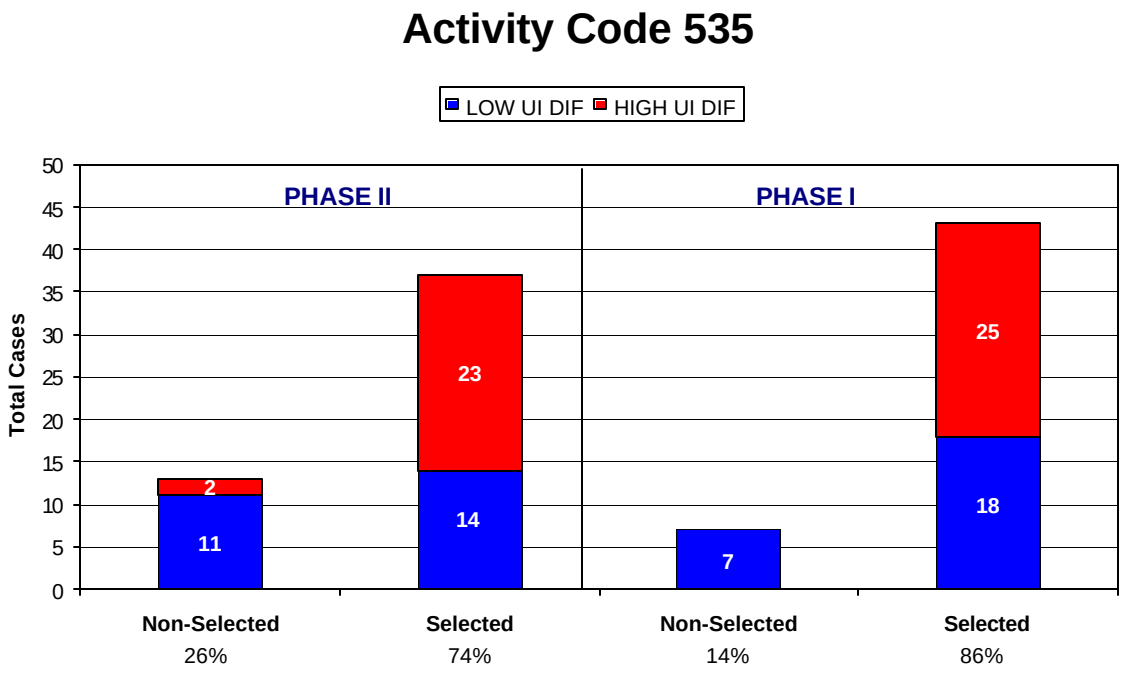


Figure 7
SELECTED AND NON-SELECTED RETURNS BY HIGH AND LOW UI DIF SCORES:

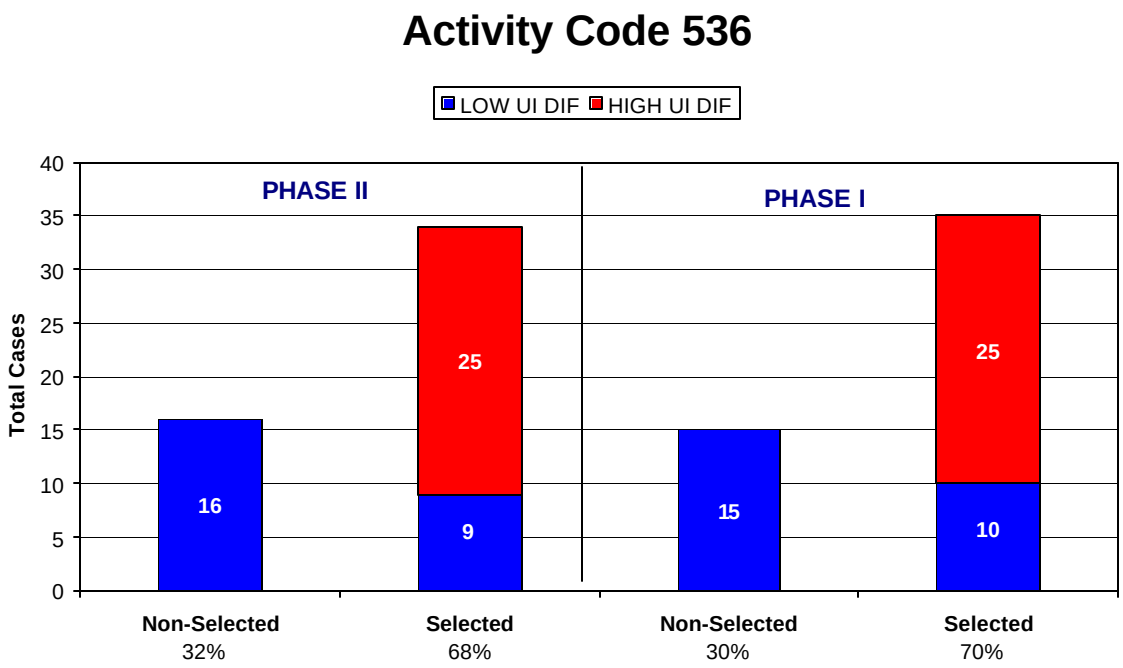
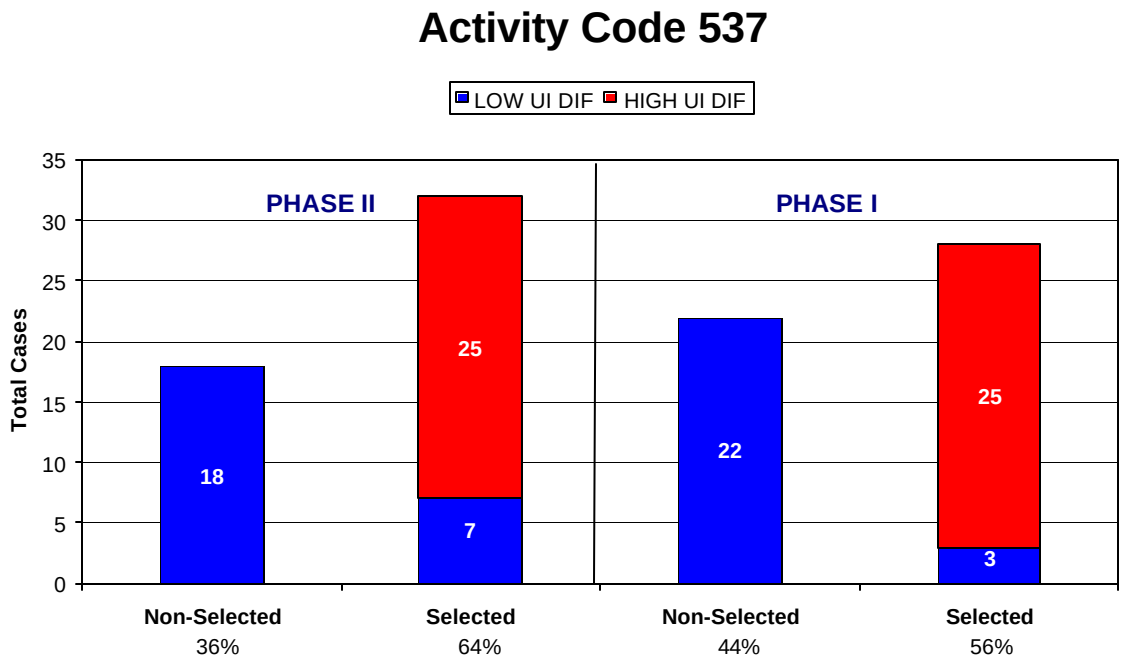


Figure 8
SELECTED AND NON-SELECTED RETURNS BY HIGH AND LOW UI DIF SCORES:



Figures 9 and 10 show the counts and percentages of selected and non-selected returns by high and low UI DIF scores over the two farming activity codes 538 and 539, for both phases. While about ½ of the returns were selected as suspicious, again, about ¾ of the selected returns were high scored.

FIGURE 9
SELECTED AND NON-SELECTED RETURNS BY HIGH AND LOW UI DIF SCORES:

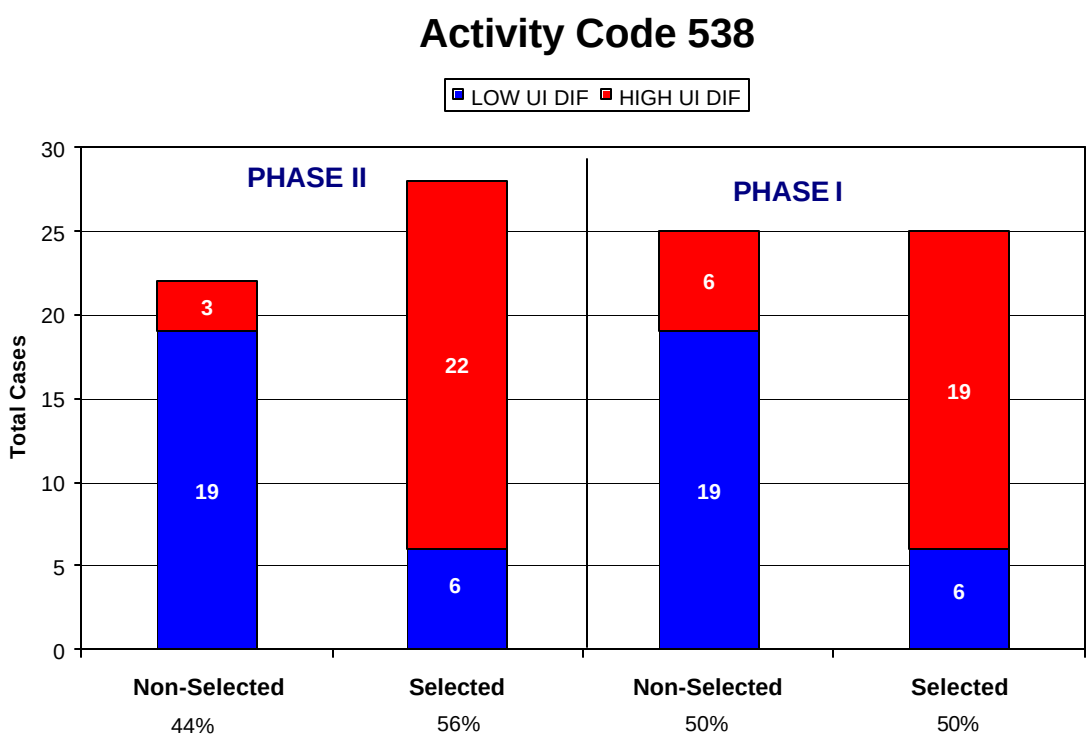
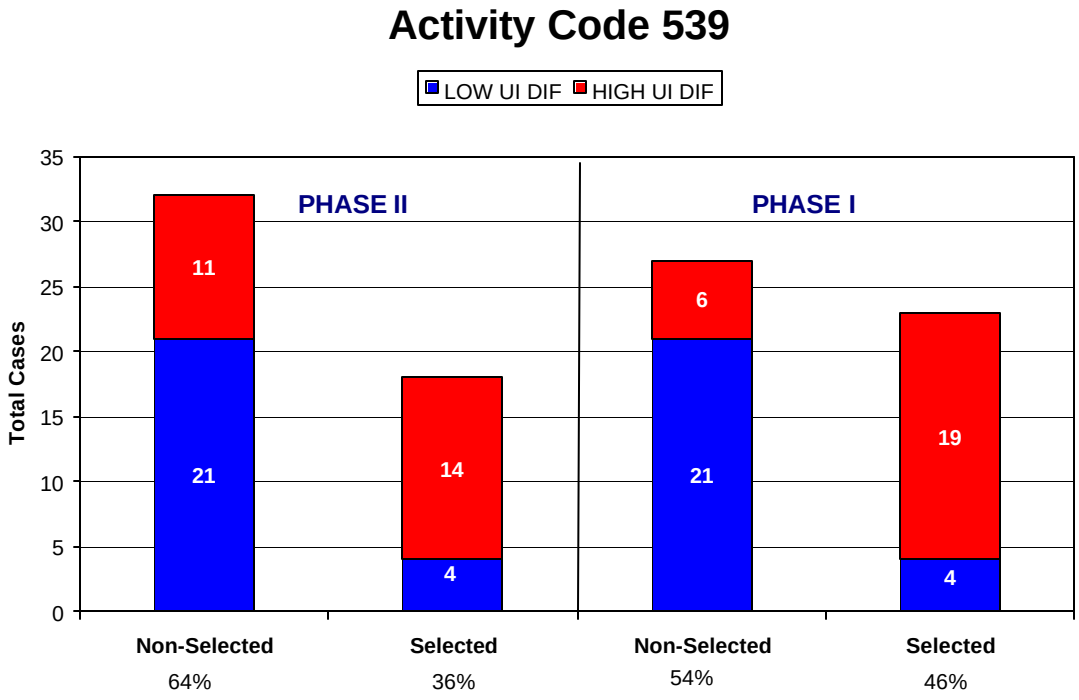


FIGURE 10
SELECTED AND NON-SELECTED RETURNS BY HIGH AND LOW UI DIF SCORES:



In phase I, the null hypothesis between scores and selections was rejected in all eight activity codes, making the results statistically significant. In Phase II, the null hypothesis was rejected in all activity codes except for activity code 534. It was retained in that activity code because the majority classifiers selection were only 2 returns. Both of the selected returns were high UI DIF scored. No low UI DIF score returns were selected.

Appendix C, D, E and F contain Tables 1,2,3 and 4 respectively that show the 2X2 tables along with counts for each cell, the selection rates for each of the eight Activity Codes for Phases I and II and the computed Fisher's value and Odds Ratio calculations. Again, these tables are computed by majority.

Figure 11 below show the ratios for the Cross-product ratio and the Log-odds ratio.

FIGURE 11
Cross-product Ratio and Log-odds Ratio - Phase II and Phase I:

	PHASE II		PHASE I	
Activity Code	**Cross-Product Ratio**	**Log-Odds Ratio**	**Cross-Product Ratio**	**Log-Odds Ratio**
532	6.25	1.83	23.08	3.14
533	7.89	2.07	16.67	2.81
534	N/A	N/A	7.89	2.07
535	9.04	2.20	9.72	2.27
536	44.44	3.79	37.5	3.62
537	64.29	4.16	183.33	5.21
538	23.22	3.15	10.03	2.31
539	6.68	1.90	16.63	2.81

The table shows what the level of association is for each activity code by phase. There is a wide range of association; however, the log-odds ratio shows us that there is a positive association in the right direction for every activity code in which there is an association.

Activity code 537 shows the strongest level of association across both phases. The high strength is reflected in the cross-product ratio. The cross-product ratio has moved far from "1" and the log-odds ratio is shown in the right direction from 0. The second strongest association is seen in activity code 536. Again, the numbers reflected for the cross-product ratio is far away from one.

The cross-product ratio increases in some activity codes and decreases in other activity codes when compared to each other by phase. The change is accounted for by the change in the actual number of cases that are selected in each phase. For example in activity code 532, the ratio decreases even though there is still some strength in the relationship. The actual number of selections decreases from phase I to phase II; thereby, causing the ratio to decrease. The selections are still high UI DIF scored which maintains the log-odds ratio going in the right direction.

There were no ratios for phase II activity code 534, since there was not a relationship shown by the Fisher's computation.

The essence of these analyses is that the classifiers generally selected high UI DIF scored returns at a higher rate than they selected low UI DIF scored returns and there is association between the UI DIF scores and the suspicion of unreported income.

Observations of the Analysis by Classifier by Activity Code

While the collective results were quite comparable between the two phases, there were individual differences by classifier in their responses. While the two phases were conducted separately and independently, many of the selections by the classifiers changed between the two phases. This was logically expected, since additional explanatory material was available. Cumulatively, over the 8 activity codes, 43 percent of the returns were not selected in either of the phases. But, of the 57 percent of the returns that were selected, just slightly more than ½ (54 percent) were selected in both trials. Slightly more than ¼ of the returns were selected only in phase I, and slightly more than 1/5 of the selected returns were selected only in phase II. The summary of classifier selections by activity code is shown in Appendix H, Table 6. In some circumstances, the additional data apparently clarified questionable areas, while, in others, it apparently raised new questions.

The rates at which individual classifiers selected returns for suspicion of unreported income varied considerably among the group. The consistency rates also varied by activity code. In the lower activity codes, consistency was low with classifiers switching their selections between the two phases; that is, they picked one particular return in phase I but did not pick it in phase II, or vice versa. Classifiers were more consistent in their selections in the higher activity codes, but they did select many returns from the low UI DIF scored sets. The rates at which the classifiers were consistent between the two phases are detailed by activity code in Appendix I.

Differences in Selections among Classifiers

In an effort to understand why returns might have elicited different responses, various individual returns were sampled at random. We looked at the returns that had a significant change in the number of selections by each individual classifier. The MACS print and the checksheets were reviewed to make an observation of why a classifier selected the return in one of the phases and then did not select it in the other phase. As stated the change occurred in selections made in phase I and not phase II and also in selections made in phase II and not in phase I.

The IRP document in the file was observed to be the item that the classifiers used the most as the basis for the select or non-select of each return. The IRP document in the file for phase II was often used to clear up questions. For some classifiers, the IRP raised new questions in phase II, thereby causing the selection of a return in phase II that we noted had been accepted as filed in phase I.

Other reasons that were provided for inconsistencies in the selections included: Missing IRP items that classifiers felt raised other questions about the possibility of inaccurate income reporting

Negative cash flow on the returns that led the classifier to believe that there could be unreported income

The appearance of related entities that were not reflected on the original return

Large Schedule E expenses that one classifier felt was the deduction of personal living expenses but another classifier felt was possibly disguising unreported income

Questions of ownership of property and whether the taxpayer could actually afford it.

Since there are separate activity codes for farm businesses, the classifiers suggested that examiners with farm-audit experience should further classify the high UI DIF scores for the farm classes 538 and 539. Some of the classifiers were uncomfortable evaluating these returns.

No-Change Rates

The purpose for conducting this study was that we wanted to be able to mimic—and, if we were successful, supplement or possibly replace—the existing operational method of identification of likely noncompliance with a tool that has been developed. That is, we want to be able to show that there is a sufficiently positive relationship between high and low UI DIF scores, and whether a return is selected by classifiers for suspicion of underreporting of income or is accepted as filed? If we succeed, then we would presumably want to use the tool.

Using the 2x2 table format we've used all along, the scoring outcomes (the results of using the tool) and the classifiers' determinations (the existing process) are consistent for observations located in the northeast and southwest cells. The results of this study show that the scores and the classifiers were consistent approximately 2/3 of the time. It is the other two corners that are addressed here. Observations found in these cells demonstrate inconsistency between scoring and classification; however, there are distinct differences between the logical consequences associated with them. In the discussion that follows, we must remember that we are proceeding on the assumption that operationally we are using the tool, the UI DIF, as the primary method of assessing apparent return accuracy. The decision made through the use of the tool determines what, if anything, occurs thereafter.

The southeast corner represents perceived compliance (per the tool). Thus, IRS operations will dictate that no further action should be taken. This erroneous outcome is described as a "false negative." The cost of false negatives is the income-underreporting component of the tax gap and can only be estimated.

The northwest corner contains the "false positive" outcomes; that is, the tool suggested that the taxpayers were noncompliant when, in fact, that was untrue. These errors generate some IRS operational action, even though inappropriate. Actual costs are incurred, which are defined as "burden." They fall on both the taxpayer and the IRS. There is also an unmeasureable—although generally agreed upon as being high—emotional burden incurred by the taxpayer. In the IRS, the rate at which taxpayers are unnecessarily audited is referred to as the “no-change” rate.

Administrators can and do make decisions that impact the costs of the false negative and positive outcomes. But, adjusting the process in order to minimize (at least, to decrease) the costs associated with false negatives automatically causes the costs associated with false positives to increase, and vice versa, unless additional information is obtained. Just as in the justice system, the IRS desires to minimize the costs associated with false positives.

If we believe that there is some relatively stable rate of noncompliance in the population of interest, we can project that rate onto whatever sample we may be interested in at any particular moment. In our study, we wanted to believe *a priori* that half of the subjects were noncompliant. The existing process that was applied (the classifiers) determined a noticeably different mixture, between activity codes.

All of the returns had received UI DIF scores. If we look only at the high-scored returns, we find again that there are differences between the actual (however we might determine this operationally) rates of noncompliance and compliance among the taxpayer types. Again, the northwest corner contains the "false positive" outcomes. These are compared to the total count of returns identified as potentially noncompliant by the tool to calculate the no-change rates.

Recently, the no-change rate for DIF-related IRS examinations has been publicized to be approximately 25%.

Traditional DIF Scores and UI DIF Scores

A question that often arises is whether or not the high UI DIF scored returns are also the returns that have high traditional DIF scores. The UI DIF development is very similar to the traditional DIF development. Only the criterion variable differs. Changes to non-IRP reported income were used instead of tax increases. The TCMP data used for formula development do not provide an opportunity to directly and independently determine if changes to reported income ultimately resulted in net increases in tax. In the 400-return random sample used in this study, 200 were purposely randomly sampled from the upper 2 percent of the UI DIF scored returns (25 from each of the 8 examination classes), and 200 were sampled from the lower 50 percent (25 from each of the 8 examination classes). Of the high-scored returns, 39 had DIF scores of at least 400, and, of the low-scored returns, 7 had DIF scores of at least 400. The distributions are shown in Appendix G, Table 5.

Less than 1 of 10 returns is both high DIF and high UI DIF scored. Nearly ¾ of these high DIF and high UI DIF scored returns were selected, either in phase I only or in both phases. None of the high DIF, low UI DIF scored returns was selected in phase II. No high traditional DIF scored returns were selected in only phase II.

Conclusions

The UI DIF scores are a very effective operational workload selection system because they mimicked the classifiers' decisions approximately 2/3 of the time overall. High UI DIF scored returns were selected for the likelihood of unreported income and low UI DIF scored returns were accepted as filed.

The UI DIF selected returns were not the same returns that would have been selected by traditional DIF scores.

UI DIF is a very good return ranking system for all types of returns.

UI DIF is a very good workload selection tool with classifiers.

In activity codes 535, 536 and 537, classifiers selected over 50% of the returns in phase I and 60% in phase II of which almost 100% of the high UI DIF scores were selected in both phases.

In activity code 538, classifiers selected 50% of the returns in phase I and 56% in phase II of which 76% in phase I and 79% in phase II had high UI DIF scores.

In activity code 539, classifiers selected 46% of the returns in phase I and 36% in phase II of which 83% in phase I and 78% in phase II had high UI DIF scores.

In activity codes 532, 533 and 534, classifiers only selected 24%, 20% & 12%, respectively of the returns in phase I and 10%, 12% & 4% in phase II; however of those returns selected, 100% had high UI DIF scores in both phases. No low UI DIF returns were selected in either phase.

UI DIF is a very good predictor of non-compliance.

Recommendations

Use UI DIF in conjunction with human classification for a return workload selection system.

Utilize UI DIF to select a predetermined portion of SB/SE compliance and W&I audit work because it selects different returns than our traditional DIF system.

Audit results from these examinations should be compared to results achieved by other workload selections (i.e. DIF).

SB/SE and W&I should utilize UI DIF along with other selection criteria for outreach.

Conduct a study of the compliance classification process because of some observed variability in the returns selected and accepted by classifiers in the two study phases.

-Study should focus on the effects of examiner experience and case building.

UI DIF should be added as a new measure on the CRIS system for predicted income change.

UI DIF should be added as a new measure for MACS.

The UI DIF formulas should be redeveloped when the National Research Program is completed. This study is expected to provide compliance data that are comparable to the TCMP studies from which the original UI DIF formulas have been produced.

APPENDIX A

Exhibit A: Instructions to the classifiers for Phase I

January 28, 2002

Welcome!

We are very pleased to have you as a participant in this very important research study. You were selected because of your experience and expertise in detecting unreported income.

Introductions: Lou Ann Sandoval, SB/SE Research Denver
Christine Rivera, Centralized Workload Selection and Delivery

Background – Estimates of the gross tax gap (the difference between what is unreported voluntarily and what should have been reported) invariably identify unreported income by sole proprietors and informal suppliers as the single largest component of the tax gap. Nearly $60 billion per year in tax revenue is lost. Detecting such noncompliance is difficult and often not cost effective using traditional methods. Various efforts have been undertaken by the Service over the years to address unreported income. The best known are the information document matching programs where income to taxpayers is reported by third parties (W-2, 1099).

Purpose of this study – To test the potential of a method for identifying the likelihood of unreported income on activity codes 532 – 539.

How this study is conducted – You will review 50 three-year return facsimiles generated by MACS, from eight activity codes (532-539). There are fewer transcribed line items than on an actual return and there will be no additional documents such as IRP for reference. You are to look at each MACS print and use your best judgment to determine whether or not you think there is a **likelihood** of unreported income. You will annotate on the scorecard a "yes, return should be examined for unreported income" or "no, return should not be examined for unreported income. You will only consider the **likelihood** of unreported income. Do not classify the return for other issues besides income. Feel free to put any comments or notes on the prints themselves for the scorecard. You each have your own set of copies. They do not have to be shared with others

You will each start with one activity code. When you have finished that AC, bring the stack with the scorecard to Lou Ann or Chris and then you will be given another AC to look at.

Definition of Unreported Income – **only that income which the taxpayer neglects to show anywhere on the return.** It may include such items as dividends received and not reported, proceeds from a sale of stock, income by a spouse who might be a direct seller, gross receipts from a business, etc. Some unreported expense items that were also missing from the return might legitimately offset income and there may be no tax consequence. Income-offsetting items such as undocumented rental or business expenses, incorrect depreciation computations, inflated costs of goods sold, etc are **NOT** considered unreported income.

Thank you for your assistance!

Exhibit B: Instructions to the classifiers for Phase II

AUSTIN PILOT - PHASE II - CLASSIFICATION WITH CASE BUILT FILES
February 21 - March 8, 2002

Purpose of this study: To test the potential of a method for identifying the likelihood of unreported income and other issues on Activity Codes 532-539 (using case built materials).

How this study is conducted: You will review 400 returns for eight Activity Codes 532-539 using the following case built materials:

- Classification Check sheet
- Checklist identifying all issues including income and the likelihood of unreported income
- Original Return/ELF Return/MACS Print (dependent upon availability)
- IDRS (Screen prints based on specific command codes - financial and t/p characteristics)
- Choice Point (Real estate and vehicle ownership)
- CBRS – dependent upon availability (cash transactions greater than $10,000)

Additional Information:

- For this pilot, it is important to identify all potential issues (do not stop at the first potential issue identified)

- We are not testing how quickly the returns are classified. It is important that time is allocated towards reviewing the case built materials to assist in the identification of potential issues.

- Please document on the Classification Sheet:
  - Check whether the return should be Selected or Accepted by checking the appropriate box.
  - If a potential income issue has been identified, check the "Examine for Unreported Income" box
  - All other issues should be documented on the Classification Checksheet
    - Description of the potential issue(s) identified
    - Specific case built material that provided the background to the potential issue
    - Rate the case built materials in terms of value in providing information towards identifying potential issue(s)

To preserve the integrity of this Research Study, please do not discuss the returns or your comments/final determinations with the other participants.

Appendix B
Exhibit C: Data Capturing Checksheet for Phase I

Appendix B

~~Sample Scorecard~~

Unreported Income Study (Phase 1)

Activity Code:
Classifier:

Should The Return Be Examined For Unreported Income?

Seq#		TIN		YES		NO
1		xxxxxxxxx				
2		xxxxxxxxx				
3		xxxxxxxxx				
4		xxxxxxxxx				
5		xxxxxxxxx				
6		xxxxxxxxx				
7		xxxxxxxxx				
8		xxxxxxxxx				
9		xxxxxxxxx				
10		xxxxxxxxx				
11		xxxxxxxxx				
12		xxxxxxxxx				
13		xxxxxxxxx				
14		xxxxxxxxx				
15		xxxxxxxxx				
16		xxxxxxxxx				
17		xxxxxxxxx				
18		xxxxxxxxx				
19		xxxxxxxxx				
20		xxxxxxxxx				
21		xxxxxxxxx				
22		xxxxxxxxx				
23		xxxxxxxxx				
24		xxxxxxxxx				
25		xxxxxxxxx				
26		xxxxxxxxx				
27		xxxxxxxxx				
28		xxxxxxxxx				
29		xxxxxxxxx				
30		xxxxxxxxx				
31		xxxxxxxxx				
32		xxxxxxxxx				
33		xxxxxxxxx				
34		xxxxxxxxx				
35		xxxxxxxxx				
36		xxxxxxxxx				
37		xxxxxxxxx				
38		xxxxxxxxx				
39		xxxxxxxxx				
40		xxxxxxxxx				
41		xxxxxxxxx				
42		xxxxxxxxx				
43		xxxxxxxxx				
44		xxxxxxxxx				
45		xxxxxxxxx				
46		xxxxxxxxx				
47		xxxxxxxxx				
48		xxxxxxxxx				
49		xxxxxxxxx				
50		xxxxxxxxx				

APPENDIX B
Exhibit D: Data Capturing Checksheet for Phase II

INTERNAL REVENUE SERVICE

Case Building Classification Checklist Time to Classify: _____

53X-XX XXX-XX-XXXX ☐ Examine for Unreported Income?	Results of Classification: ☐ 1. Accepted as Filed ☐ 2. Selected for Examination

Issue #	X	If Issue(s) are questionable, "X" left column:	Remarks:
04		Filing Status (Married Filing Separately)	
05		Filing Status (Head of Household)	
		EXEMPTIONS:	
06		Dependents who live with TP	
07		Other Dependents	
		INCOME:	
08		Income W2/1099	
09		Other Income	
10		IRMF	
		ADJUSTMENTS TO INCOME:	
11		Individual Retirement Arrangements	
12		Alimony Payments	
13		Self Employment Health Insurance	
		ITEMIZED DEDUCTIONS – SCHEDULE A:	
14		Medical and Dental Expenses	
47		Medical Savings Account	
15		State and Local Income Taxes/State Tax Refund	
16		Real Estate and Personal Property Taxes	
17		Interest Expense	
18		Contributions	
19		Casualty and Theft Losses	
20		Moving Expenses	
21		Miscellaneous Deductions (other than EBE)	
46		Student Loan Interest	
		EMPLOYEE BUSINESS EXPENSE:	
22		All Employee Business Expenses	
23		Automobile Expenses	
24		Entertainment, Meals, Gifts, and Other Expenses	
25		Travel, Lodging and Other Expenses	
30		Business Use of Home	
31		Education Expenses	
		GAINS AND LOSSES:	
26		Bad Debts	
27		Stock Sales	
28		Schedule D, Sales of Real and Personal Property	
29		Sale of Personal Residence	
		OTHER TAXES:	
32		Alternative Minimum Tax	
36		Self Employment Tax	
37		Other Taxes	
		TAX CREDITS:	
90		Child and Dependent Care Credit	
91		Foreign Tax Credit	
92		Earned Income Credit	
93		Child Tax Credit	
94		Education Credit	
95		Adoption Credit	

Other Issues (Insert issue exactly as shown on return):

Classifier Name:	Date Classified:	Reviewed: ☐ With ☐ Without Screener	Date Reviewed:

	"X" if questionable:									Issue(s) on Sch C, E or F:	Remarks:
ssue #	C 1	C 2	C 3	E 1	F 1	F 2					
40										Rental Income (Schedule E)	
50										Gross Receipts/Income	
52										All Expenses	
53										Advertising	
54										Bad Debts from Sales or Services	
55										Car & Truck Expenses	
41										Cleaning and Maintenance	
56										Commissions, Cost of Labor, Labor Hired	
57										Cost of Goods Sold	
58										Cost of Livestock/Other Items Bought for Resale	
59										Depreciatory/Section 179	
42										Employee Benefit Program	
43										Freight & Trucking	
44										Feed Purchased	
45										Seed & Plants Purchased	
51										Fertilizer, Lime, Chemicals	
62										Gas, Oil, & Fuel	
63										Insurance	
64										Depletion	
65										Interest Expense	
66										Legal and Professional Expense	
67										Custom Hire/Machine Work	
68										Office Expenses	
69										Office in the Home	
70										Pension/Profit Sharing Plans	
71										Rent or Lease	
72										Repairs & Maintenance	
73										Wages	
74										Self Employment Tax	
75										Supplies	
76										Taxes	
77										Travel, Meals & Entertainment	
78										Utilities	
79										Veterinary, Breeding & Entertainment	
80										IRC 183 Business History	
81										Other Expenses	
82										Passive Activity Losses	
60										Form 4797, State of Business Property	

Other Issues (Insert issue exactly as shown on return)

Evaluation of Case Building Documents

Case Building Document Type:	1	2	3	NP
MACS:				
CBRS:				
CHOICEPOINT:				
IDRS:				
IRP:				
Dependent Data Base				

Key for rating benefit of casebuilding documents in the classification process:

1 = Valuable
2 = Somewhat valuable
3 = Not valuable
NP = Document was not present in case file

Appendix C
Table 1: Summary of Majority Classifier Counts for each cell, selection rates and Statistic calculations Phase I and II - Activity Codes 532 and 533

Activity Code 532 -Phase II							**Activity Code 532 -Phase I**						
Fisher's Exact Calculation							**Fisher's Exact Calculation**						
	Classifier Opinion							Classifier Opinion					
UI-DIF	Non-Selected	Selected			**Fisher's Exact Test**	0.0251	UI-DIF	Non-Selected	Selected			**Fisher's Exact Test**	0.0000
High	20	5	25		- reject Null Hypothesis		High	13	12	25		- reject Null Hypothesis	
Low	25	0	25				Low	25	0	25			
	45	5	50					38	12	50			
Odds Ratio Calculation							**Odds Ratio Calculation**						
Set cell to one if cell is zero in Fisher's Test							Set cell to one if cell is zero in Fisher's Test						
UI DIF	Non-Selected	Selected					UI DIF	Non-Selected	Selected				
High	20	5	25		**Cross-Product Ratio**	6.25	High	13	12	25		**Cross-Product Ratio**	23.08
Low	25	1	26		**Log-Odds Ratio**	1.83	Low	25	1	26		**Log-Odds Ratio**	3.14
	45	6	51					38	13	51			
Activity Code 533 -Phase II							**Activity Code 533 -Phase I**						
Fisher's Exact Calculation							**Fisher's Exact Calculation**						
	Classifier Opinion							Classifier Opinion					
UI-DIF	Non-Selected	Selected			**Fisher's Exact Test**	0.0111	UI-DIF	Non-Selected	Selected			**Fisher's Exact Test**	0.0003
High	19	6	25		- reject Null Hypothesis		High	15	10	25		- reject Null Hypothesis	
Low	25	0	25				Low	25	0	25			
	44	6	50					40	10	50			
Odds Ratio Calculation							**Odds Ratio Calculation**						
Set cell to one if cell is zero in Fisher's Test							Set cell to one if cell is zero in Fisher's Test						
UI DIF	Non-Selected	Selected					UI DIF	Non-Selected	Selected				
High	19	6	25		**Cross-Product Ratio**	7.89	High	15	10	25		**Cross-Product Ratio**	16.67
Low	25	1	26		**Log-Odds Ratio**	2.07	Low	25	1	26		**Log-Odds Ratio**	2.81
	44	7	51					40	11	51			

Appendix D
Table 2: Summary of Majority Classifier Counts for each cell, selection rates and Statistic calculations Phase I and II - Activity Codes 534 and 535

Activity Code 534 -Phase II								**Activity Code 534 -Phase I**						
Fisher's Exact Calculation								**Fisher's Exact Calculation**						
	Classifier Opinion								Classifier Opinion					
UI-DIF	Non-Selected	Selected			**Fisher's Exact Test**	0.2449		UI-DIF	Non-Selected	Selected			**Fisher's Exact Test**	0.0111
High	23	2	25		-retain Null			High	19	6	25		- reject Null Hypothesis	
Low	25	0	25					Low	25	0	25			
	48	2	50						44	6	50			
Odds Ratio Calculation								**Odds Ratio Calculation**						
Set cell to one if cell is zero in Fisher's Test								Set cell to one if cell is zero in Fisher's Test						
UI DIF	Non-Selected	Selected						UI DIF	Non-Selected	Selected				
High	23	2	25					High	19	6	25		**Cross-Product Ratio**	7.89
Low	25	1	26					Low	25	1	26		**Log-Odds Ratio**	2.07
	48	3	51						44	7	51			
Activity Code 535 -Phase II								**Activity Code 535 -Phase I**						
Fisher's Exact Calculation								**Fisher's Exact Calculation**						
	Classifier Opinion								Classifier Opinion					
UI-DIF	Non-Selected	Selected			**Fisher's Exact Test**	0.0038		UI-DIF	Non-Selected	Selected			**Fisher's Exact Test**	0.0048
High	2	23	25		- reject Null Hypothesis			High	0	25	25		- reject Null Hypothesis	
Low	11	14	25					Low	7	18	25			
	13	37	50						7	43	50			
Odds Ratio Calculation								**Odds Ratio Calculation**						
Set cell to one if cell is zero in Fisher's Test								Set cell to one if cell is zero in Fisher's Test						
UI DIF	Non-Selected	Selected						UI DIF	Non-Selected	Selected				
High	2	23	25		**Cross-Product Ratio**	9.04		High	1	25	26		**Cross-Product Ratio**	9.72
Low	11	14	25		**Log-Odds Ratio**	2.20		Low	7	18	25		**Log-Odds Ratio**	2.27
	13	37	50						8	43	51			

Appendix E
Table 3: Summary of Majority Classifier Counts for each cell, selection rates and Statistic calculations Phase I and II - Activity Codes 536 and 537

Activity Code 536 -Phase II

Fisher's Exact Calculation

	Classifier Opinion		
UI-DIF	Non-Selected	Selected	
High	0	25	25
Low	16	9	25
	16	34	50

Fisher's Exact Test 0.0000
- reject Null Hypothesis

Odds Ratio Calculation

Set cell to one if cell is zero in Fisher's Test

UI DIF	Non-Selected	Selected	
High	1	25	26
Low	16	9	25
	17	34	51

Cross-Product Ratio 44.44
Log-Odds Ratio 3.79

Activity Code 536 -Phase I

Fisher's Exact Calculation

	Classifier Opinion		
UI-DIF	Non-Selected	Selected	
High	0	25	25
Low	15	10	25
	15	35	50

Fisher's Exact Test 0.0000
- reject Null Hypothesis

Odds Ratio Calculation

Set cell to one if cell is zero in Fisher's Test

UI DIF	Non-Selected	Selected	
High	1	25	26
Low	15	10	25
	16	35	51

Cross-Product Ratio 37.50
Log-Odds Ratio 3.62

Activity Code 537 -Phase II

Fisher's Exact Calculation

	Classifier Opinion		
UI-DIF	Non-Selected	Selected	
High	0	25	25
Low	18	7	25
	18	32	50

Fisher's Exact Test 0.0000
- reject Null Hypothesis

Odds Ratio Calculation

Set cell to one if cell is zero in Fisher's Test

UI DIF	Non-Selected	Selected	
High	1	25	26
Low	18	7	25
	19	32	51

Cross-Product Ratio 64.29
Log-Odds Ratio 4.16

Activity Code 537 -Phase I

Fisher's Exact Calculation

	Classifier Opinion		
UI-DIF	Non-Selected	Selected	
High	0	25	25
Low	22	3	25
	22	28	50

Fisher's Exact Test 0.0000
- reject Null Hypothesis

Odds Ratio Calculation

Set cell to one if cell is zero in Fisher's Test

UI DIF	Non-Selected	Selected	
High	1	25	26
Low	22	3	25
	23	28	51

Cross-Product Ratio 183.33
Log-Odds Ratio 5.21

Appendix F
Table 4: Summary of Majority Classifier Counts for each cell, selection rates and Statistic calculations Phase I and II - Activity Codes 538 and 539

Activity Code 538 -Phase II								**Activity Code 538 -Phase I**						
Fisher's Exact Calculation								**Fisher's Exact Calculation**						
	Classifier Opinion								Classifier Opinion					
UI-DIF	Non-Selected	Selected			**Fisher's Exact Test**	0.0000		UI-DIF	Non-Selected	Selected			**Fisher's Exact Test**	0.0002
High	3	22	25		- reject Null Hypothesis			High	6	19	25		- reject Null Hypothesis	
Low	19	6	25					Low	19	6	25			
	22	28	50						25	25	50			
Odds Ratio Calculation								**Odds Ratio Calculation**						
Set cell to one if cell is zero in Fisher's Test								Set cell to one if cell is zero in Fisher's Test						
UI DIF	Non-Selected	Selected						UI DIF	Non-Selected	Selected				
High	3	22	25		**Cross-Product Ratio**	23.22		High	6	19	25		**Cross-Product Ratio**	10.03
Low	19	6	25		**Log-Odds Ratio**	3.15		Low	19	6	25		**Log-Odds Ratio**	2.31
	22	28	50						25	25	50			
Activity Code 539 -Phase II								**Activity Code 539 -Phase I**						
Fisher's Exact Calculation								**Fisher's Exact Calculation**						
	Classifier Opinion								Classifier Opinion					
UI-DIF	Non-Selected	Selected			**Fisher's Exact Test**	0.0031		UI-DIF	Non-Selected	Selected			**Fisher's Exact Test**	0.0000
High	11	14	25		- reject Null Hypothesis			High	6	19	25		- reject Null Hypothesis	
Low	21	4	25					Low	21	4	25			
	32	18	50						27	23	50			
Odds Ratio Calculation								**Odds Ratio Calculation**						
Set cell to one if cell is zero in Fisher's Test								Set cell to one if cell is zero in Fisher's Test						
UI DIF	Non-Selected	Selected						UI DIF	Non-Selected	Selected				
High	11	14	25		**Cross-Product Ratio**	6.68		High	6	19	25		**Cross-Product Ratio**	16.63
Low	21	4	25		**Log-Odds Ratio**	1.90		Low	21	4	25		**Log-Odds Ratio**	2.81
	32	18	50						27	23	50			

Appendix G
Table 5: High DIF Scored Returns

Activity Code		Count	Selected in Phase I Only	Selected in Both Phases	Selected in Neither Phase
532	High UI DIF	3	2	1	
	Low UI DIF	0			
	Total	3	2	1	0
533	High UI DIF	5	2	1	2
	Low UI DIF	1			1
	Total	6	2	1	3
534	High UI DIF	1	1		
	Low UI DIF	0			
	Total	1	1	0	0
535	High UI DIF	12	2	10	
	Low UI DIF	0			
	Total	12	2	10	0
536	High UI DIF	6		6	
	Low UI DIF	3	2		1
	Total	9	2	6	1
537	High UI DIF	0			
	Low UI DIF	1			1
	Total	1	0	0	1
538	High UI DIF	3	1		2
	Low UI DIF	2			2
	Total	5	1	0	4
539	High UI DIF	9	3	1	5
	Low UI DIF	0			
	Total	9	3	1	5
Combined	High UI DIF	39	11	19	9
	Low UI DIF	7	2	0	5
	Total	46	13	19	14

Appendix H
Table 6: Summary of Classifier Selections for Phase I and II

Activity Code	High/Low	Phase I Averages	Phase II Averages	Ph. I Select	Ph. II Select	I Hi/Lo Ratio	II Hi/Lo Ratio	Select I & II	Select I Only	Select II Only	Select Neither	Check Sum	Select I & II	Select I Only	Select II Only	Select Neither
532	H	11.82	8.18	47.27%	32.73%	9.29	4.09	63	67	27	118	275	22.91%	24.36%	9.82%	42.91%
	L	1.27	2.00	5.09%	8.00%			2	12	20	241	275	0.73%	4.36%	7.27%	87.64%
		13.09	10.18	26.18%	20.36%			65	79	47	359	550	11.82%	14.36%	8.55%	65.27%
533	H	10.91	8.09	43.64%	32.36%	12.00	7.42	49	71	40	115	275	17.82%	25.82%	14.55%	41.82%
	L	0.91	1.09	3.64%	4.36%			0	10	12	253	275	0.00%	3.64%	4.36%	92.00%
		11.82	9.18	23.64%	18.36%			49	81	52	368	550	8.91%	14.73%	9.45%	66.91%
534	H	7.64	7.64	30.55%	30.55%	4.94	6.46	35	49	49	142	275	12.73%	17.82%	17.82%	51.64%
	L	1.55	1.18	6.18%	4.73%			1	16	12	246	275	0.36%	5.82%	4.36%	89.45%
		9.18	8.82	18.36%	17.64%			36	65	61	388	550	6.55%	11.82%	11.09%	70.55%
535	H	23.00	19.82	92.00%	79.27%	1.35	1.46	209	44	9	13	275	76.00%	16.00%	3.27%	4.73%
	L	17.09	13.55	68.36%	54.18%			131	57	18	69	275	47.64%	20.73%	6.55%	25.09%
		40.09	33.36	80.18%	66.73%			340	101	27	82	550	61.82%	18.36%	4.91%	14.91%
536	H	22.18	22.00	88.73%	88.00%	2.28	2.22	220	24	22	9	275	80.00%	8.73%	8.00%	3.27%
	L	9.73	9.91	38.91%	39.64%			69	38	40	128	275	25.09%	13.82%	14.55%	46.55%
		31.91	31.91	63.82%	63.82%			289	62	62	137	550	52.55%	11.27%	11.27%	24.91%
537	H	21.09	20.82	84.36%	83.27%	3.68	2.83	196	36	33	10	275	71.27%	13.09%	12.00%	3.64%
	L	5.73	7.36	22.91%	29.45%			33	30	48	164	275	12.00%	10.91%	17.45%	59.64%
		26.82	28.18	53.64%	56.36%			229	66	81	174	550	41.64%	12.00%	14.73%	31.64%
538	H	16.73	16.36	66.91%	65.45%	2.19	1.70	132	52	48	43	275	48.00%	18.91%	17.45%	15.64%
	L	7.64	9.64	30.55%	38.55%			53	31	53	138	275	19.27%	11.27%	19.27%	50.18%
		24.36	26.00	48.73%	52.00%			185	83	101	181	550	33.64%	15.09%	18.36%	32.91%
539	H	14.55	14.18	58.18%	56.73%	2.11	2.00	109	51	47	68	275	39.64%	18.55%	17.09%	24.73%
	L	6.91	7.09	27.64%	28.36%			34	42	44	155	275	12.36%	15.27%	16.00%	56.36%
		21.45	21.27	42.91%	42.55%			143	93	91	223	550	26.00%	16.91%	16.55%	40.55%
Combined	H	15.99	14.64	63.95%	58.55%	2.52	2.26	1013	394	275	518	2200	46.05%	17.91%	12.50%	23.55%
	L	6.35	6.48	25.41%	25.91%			323	236	247	1394	2200	14.68%	10.73%	11.23%	63.36%
		22.34	21.11	44.68%	42.23%			1336	630	522	1912	4400	30.36%	14.32%	11.86%	43.45%
													53.70%	25.32%	20.98% of Selects	

Appendix I: Consistency Rates by Activity Code

Activity Code 532: For this activity code, each classifier selected many more of the high UI DIF 532 returns versus the low ones. For the high 532 returns, the selects among each classifier from Phase I to Phase II were not very consistent. Only one classifier was over 50% consistent in their selections between the two phases. That classifier selected the same 15 returns in both phases. The consistency rate varied from 8% to the high of 60% with the average was 23% of the time that the same returns in Phase I and Phase II for the high UI DIF 532 returns were selected. This means that only 1 out of 4 times did the same return get selected between Phase I and Phase II.

The select rate for the low UI 532 returns was low. The highest select rate by any one classifier was 7 selects in Phase II. The selections from Phase I to Phase II consistently changed. The classifiers selected different returns in Phase I from Phase II. They had a 0% consistency rate for selections by nine of the classifiers. There was only one classifier who selected the same return in both Phase I and Phase II (4% consistent); otherwise they were all different selects. One classifier had no selections of any low UI returns in either Phase I or Phase II.

Activity Code 533: Each classifier selected more of the high UI DIF 533 returns than the low UI returns. The consistency rate of the selections by each classifier was low, with most at 8% or 12%. Only one classifier had a consistency rate of over 50%. That classifier selected 14 of the 25 returns in both phases. The average consistency rate was 17% of the time did the same returns get selected in Phase I and Phase II. Most of the time, the classifiers were switching their selections. Two classifiers changed their selections over 60% of the time. They selected different returns in Phase I then they selected in Phase II. Six classifiers selected more returns in Phase I and three classifiers selected more returns in Phase II. Two classifiers selected the same number of returns in both phases; however, they were not all the returns.

The select rate for the low UI DIF 533 returns was low. The highest number of selects was 4 returns in Phase I by one classifier. Most of the classifiers only selected one or two returns with 9 of the 11 making no selections during at least one of the phases. There was a 0% consistency selection rate of the low UI returns. None of the classifiers selected the same low returns in both Phase I and Phase II. All selections changed between the two phases.

Activity Code 534: Again, each classifier selected more of the high UI DIF 534 returns than the low ones. There were no classifiers who consistently selected the same returns in both Phase I and Phase II over 50% of the time. The highest consistency rate was 32%, meaning that 1 out of 3 selects was the same in Phase I and Phase II. The average consistency rate was 13%. The average change of the selections was 36%. The classifiers changed their selections between the two phases very often. On average each classifier changed selections on 1 out of every 3 returns. One classifier changed his selections 64% of the time. That classifier had four selections in Phase I and 18 in Phase II with three of the selections being the same return in both phases.

The select rate for the low UI 534 returns was low. The highest select rate by any one classifier was 6 selects in Phase I. Most of the classifiers only selected between 1 and 5 returns. Nine of the 11 classifiers made no selection in at least one of the phases. The selections from Phase I to Phase II consistently changed. The classifiers selected different returns in Phase I from Phase II. There was a 0% consistency rate for selections by nine of the classifiers. Only one

classifier selected one return in both Phase I and Phase II (4% consistent); otherwise all the selections in Phase I and Phase II differed. One classifier had no selections of any low UI returns in either phase.

Activity Code 535: The select rate of the high UI DIF 535 returns was extremely high. One classifier selected all 25 of the high returns in both Phase I and Phase II. In Phase I, the lowest number of selections was 13 by one classifier. All other ten classifiers selected 21 or more of the returns. In Phase II, the selections remained high. The lowest select rate was 12 selections by one classifier. The remaining ten classifiers selected between 17 and 25 returns. The consistency rate for 10 of the 11 classifiers was over 64%. This meant that they selected many of the same returns in both Phase I and Phase II. Their selection rate change averaged 19%.

The select rate for the low UI 535 returns was also high. In Phase I, only one of the classifiers selected less than 10 of the returns. The selections of the other 10 classifiers ranged from 11 to 24 selections. In Phase II, the number of selections decreased for 10 of the 11 classifiers. One classifier had more selections of the low returns in Phase II. That classifier went from 11 selects in Phase I to 16 selects in Phase II, with 10 of the selects being the same for both phases. The consistency rate of the selections between Phase I and Phase II was high at above 60%.

Activity Code 536: The select rate for the high UI DIF 536 returns was high. The consistency rate of selections was also high at above 50% for 10 of the 11 classifiers. The average consistency rate was 80%. One classifier had a selection consistency rate of 44%. The classifier doubled their selections from 11 selects in Phase I to 22 selects in Phase II. One classifier selected all 25 returns in both Phase I and Phase II. The lowest select rate was the 11 in Phase I and 16 in Phase II.

The low UI 536 returns had a high select rate among some classifiers and a low select rate among others. One classifier selected only 4 returns between both phases with none of them being the same return selections. One classifier selected 21 returns in Phase I and 15 in Phase II with 13 of those being the same selections in Phase I and Phase II. The number of selections among the remaining classifiers varied. The selections in Phase I and Phase II were consistent on average 25% of the time. The classifiers selected the same return 1 out of every 4 times.

Activity Code 537: The selection rate for the high UI DIF 537 returns was high. The consistency rate of selections remained high at above 50% for 8 of the 11 classifiers, this included 4 of the classifiers with a rate of 96%. The remaining 3 classifiers were consistent at 40%, 44% & 48%, respectively. The average consistency rate was 71%. Six of the classifiers selected all 25 high returns at least during one of the phases. The lowest selection rate was 13 selects in Phase I and 15 selects in Phase II.

The low UI 537 returns were selected often by some of the classifiers. One classifier increased the selections from 1 in Phase I to 15 in Phase II. Two other classifiers also increased their selections from Phase I to Phase II. Eight classifiers decreased the number of selections they made in Phase II from their selections in Phase I. The selection consistency rate for the low returns varied from 0% consistent to 52% consistent. The average rate was 12%. The classifiers with few selections had the lower consistency rates. They did not make the same selections between the two phases.

Activity Code 538: The selection rate for the high UI DIF 538 returns remained high with the exception of one classifier who only had two selections in Phase II. One classifier increased the selections from 3 in Phase I to 23 in Phase II. Eight of 11 classifiers decreased the number of selections in Phase II and three increased their number of selections. The selection consistency rate was over 60% for five of the classifiers. They continued to select a lot of the same cases in both phases. One classifier was only consistent 8% of the time. The average consistency rate was 48%.

The classifiers selected the low UI 538 returns often. One classifier increased the selections from 0 in Phase I to 13 selections in Phase II. A total of six classifiers increased their number of selections in Phase II. The consistency rate of the low selections by each classifier was low. The classifiers picked a lot of different returns in Phase I than those in Phase II. The average consistency rate was 19%.

Activity Code 539: The select rate of the high UI DIF 539 returns was high with the exception of one classifier who only had a total of 5 selections in both Phase I and Phase II. Seven of the 11 classifiers decreased the number of selects in Phase II. Three classifiers increased their number of selections in Phase II and one classifier selected 14 returns both times; however only 7 of the returns were the same ones in both phases. The consistency rate of the selections was over 50% for five of the classifiers. The other six had a consistency rate below 36%.

The low UI 539 returns select rate varied from high to low. One classifier only selected 2 returns in Phase II while another classifier selected 17 returns in Phase I and 15 in Phase II of which 10 were the same returns. The selection consistency rate of the selections was low at 0% to 8% with the exception of one classifier who had a consistency rate of 40%, another with 28% and a third with 20%.

ABSTRACT

Estimates of the gross tax gap invariably identify unreported income by sole proprietors as the single largest component of the individual income tax gap. Detecting such noncompliance is difficult using traditional methods. This study of the unreported income discriminant function (UI DIF) scores was designed to test the consistency between whether an individual income tax return was high- or low-UI DIF scored and whether expert classifiers either selected that return for suspicion of unreported income or accepted it as filed. The high- and low-scored returns had been randomly sampled and scrambled together within the eight examination classes (activity codes 532-539) for which the formulas have been developed.

The tests conducted on activity codes 535, 536 and 537 and the farm returns, activity codes 538 and 539 suggest that the UI DIF scores, in general, would be very useful. The tests conducted on activity codes 532, 533 and 534 suggest that the UI DIF scores may not be as useful. Fisher's exact test showed a weak association or no association between the UI DIF scores and selection rates for these activity codes. The UI DIF scores identify different returns than do the traditional DIF scores. The classification results themselves have not yet been validated with examination results. This additional effort is currently underway.

KEY WORD: UI DIF, DIF, Workload Selection, Scores, Classification, high UI DIF scored, Low UI DIF scored, activity code, 532, 533, 534, 535, 536, 537, 538, 539, 2X2 table, Fisher, odds ratio

ACKNOWLEDGEMENT

The Denver field office of Small Business/Self-Employed Research expresses its gratitude to the DIF-development team in the Studies and Modeling B group in National Headquarters Research. We could not have resurfaced the UI DIF return-scoring methodology without their continuing assistance. For that we are thankful.