



An Assessment of the Asserted Increase in ‘Certainty’ Arising from the Introduction of the Tax Value Method (TVM)

Final Report

Graeme Cooper
Professor Taxation Law
Faculty of Law, University of Melbourne

&

Michael Wenzel
Fellow
Centre for Tax System Integrity, Research School of Social Sciences
Australian National University

6 June 2002

Summary

This document reports the results of a research project, “An Assessment of the Asserted Increase in ‘Certainty’ Arising from the Introduction of the Tax Value Method (TVM)”, commissioned by the Board of Taxation. The project was designed to elicit a quantifiable measure of the extent to which TVM enhances or detracts from the degree of “certainty” in the current income tax rules which define the tax base.

In summary, the results of our project show:

- there is reason to believe that TVM will lead to *less consistency* in reaching consensus on the answers to tax problems, particularly in some areas of tax law,
- there is reason to believe TVM will lead to *less confidence* in the correctness of answers taxpayers reach to their tax problems,
- there is reason to believe TVM will lead to *less confidence* that taxpayers are correctly following the steps which the law requires to determine answers to their tax problems, and
- there is reason to believe that TVM will lead taxpayers to reach *less accurate* answers to their tax problems.

There is now some evidence to assess the claim that TVM will lead to greater certainty in the tax system and the evidence, so far, is that it won't.

1. Background

This document reports to the Board of Taxation the results of the research project, “The Assessment of the Asserted Increase in ‘Certainty’ Arising from the Introduction of the Tax Value Method (TVM)”, commissioned by the Board on 14 December 2001 from the University of Melbourne in collaboration with the Centre for Tax System Integrity, Research School of Social Sciences, Australian National University. This document is the written Final Report on completion of the project for the purposes of clause B(c) of the Schedule to the agreement between the Board and the University of Melbourne.

The project was designed to elicit a quantifiable measure of the extent to which the TVM being evaluated by the Board of Taxation enhances or detracts from the degree of “certainty” expressed in the current income tax law in defining the legal concept of “taxable income.” This document explains:

- the design of the experiment and the review process to which it was subjected,
- the conduct of the experiments,
- the results of the experiments, and
- the interpretation of the data and conclusions.

Annexed to the Report are the following documents:

Experimental review documents

- Letter requesting peer review of the experiment and materials (Attachment A)
- Review from Associate Professor John Glover (Attachment B)
- Review from Professor Cameron Rider (Attachment C)
- Letter from TVM Legislative Team (Attachment D)

Experiment documents

- Advertisement for participants (Attachment E)
- Instructions to participants in the experiment (Attachment F)
- Text of presentation on TVM (Attachment G)
- Text of presentation on current law (Attachment H)
- Experimental questions in TVM format (Attachment I)
- Experimental questions in current law format (Attachment J)
- Extracts of TVM legislation used in experiments (Attachment K)
- Extracts of current legislation used in experiments (Attachment L)

Evaluative documents

- Explanation of intended answers to questions based on TVM and current law (Attachment M)
- Table of observed answers and their frequencies for TVM and current legislation (Attachment N)

2. Experimental issue

The current expression of the base of the income tax (“taxable income”) in ITAA 36 and ITAA 97 examines individual receipts and classifies them as “income” or not, and examines individual payments and losses and classifies them as “deductions” or not. The rules which classify receipts and outgoings are currently expressed in s. 6-5 and s. 8-1 ITAA 97. Each term is supplemented by various statutory formulations of the same idea and various statutory expansions. If receipts or outgoings are classified as income or expenses, certain consequences follow; if they are not, other consequences follow.

This formulation will change if Recommendation 4.1 of the Review of Business Taxation [RBT] is implemented by Government. This recommendation is to re-express the income tax base using the formulation usually referred to as “Option 2” or the “Tax Value Method.” The formulation received “in principle support” by the Government but was referred to the Board of Taxation by the Treasurer with a request that the Board examine the Recommendation further [Press Release 81/2000, Treasurer, Tax Value Method (7 August 2000)].

This project was commissioned as part of the Board’s examination of the TVM. The claim which this research project was designed to test is the assertion in the Final Report of the RBT that TVM will lead to “*reducing uncertainty* and complexity in the present system” [Review of Business Taxation, *A Tax System Redesigned* (1999) p. 156,]. ‘Certainty’ is a notion to which it is necessary to ascribe a more exact meaning for the purposes of this research because there are many plausible meanings of ‘certainty’ which the RBT might have been using.

The RBT might have been asserting that TVM will lead to any or more of these outcomes:

- greater consensus on a likely outcome,
- greater consistency in reaching an intended outcome,

- greater consistency in predicting the outcome of unforeseen events,
- greater confidence in the accuracy of results reached.

Whichever meaning of ‘certainty’ was intended by the RBT, the meaning could be tested in a variety of ways – for example, by showing:

- more people agreeing on the same answer under TVM than under current law (we call this, *objective certainty*),
- fewer people agreeing on an answer, but the dispersion of their answers is smaller (this is a measure of *objective certainty*),
- people being on average more confident in the correctness of their decisions under TVM than under current law (we call this *subjective certainty*),
- more people reaching the answer which the legislature intended (we call this *substantive accuracy*), or
- more people accurately applying the processes which the rules prescribe to reach the answer which the legislature intended (we call this *procedural accuracy*).

These tests could be used to measure ‘certainty’ in each of three plausible senses – consensus, confidence and accuracy. The experiment focussed on these three ideas as the most plausible definitions of ‘certainty’ and the most susceptible to analysis.

2.1 Principal hypotheses

In order to test the claim that the TVM would increase ‘certainty’ in tax law in these three senses, we formulated and investigated five specific hypotheses. First, we test the prediction that TVM will yield greater *objective certainty*: more people agree on the same answer under TVM than under current law. We test the hypothesis:

H1 Compared to the current law, TVM will lead to greater consistency in reaching answers to common tax problems.

Second, we test the prediction that TVM will yield greater *subjective certainty*, in that taxpayers will indicate greater confidence (a) in the correctness of their proposed answer and (b) that they have correctly followed and applied the steps required by law to determine the outcome:

H2 Compared to the current law, TVM will lead to greater subjective confidence in the correctness of answers to tax problems.

H3 Compared to the current law, TVM will lead to greater subjective confidence that the steps used to determine the answers to tax problems have been followed correctly.

2.2 Secondary hypotheses

The hypotheses just described do not include any reference to ‘reaching the right answer’ – mere consensus is the evidence of certainty. This agnosticism was a considered position because of the difficulty, even for tax experts, of reaching agreement on ‘the right answers’ to many tax problems, and it would be an important (perhaps even sufficient) achievement if TVM were able to induce more people or more often to the same answer, without also expecting people to agree on ‘the right answer.’

Nevertheless, if the claim of greater ‘certainty’ is understood to imply greater consistency in reaching the Legislature’s *intended* outcome, *accuracy* is also obviously a relevant line of inquiry for this project. Since we had already collected the data, we decided to expand our inquiry to test two subsidiary claims, whether (a) taxpayers demonstrate greater agreement with our view of the intended outcome to tax problems and (b) whether taxpayers demonstrate greater accuracy in applying the steps by which the legislature intended the outcomes to be determined. Hence, we test two subsidiary hypotheses:

H4 Compared to the current law, TVM will lead to greater frequency in reaching the intended answers to tax problems.

H5 Compared to the current law, TVM will lead to greater consistency in applying the steps intended to be applied in order to determine answers to tax problems.

Hypothesis H5 deserves some more explanation. Some elements of TVM, just as in the current law, require taxpayers to follow quite specific but rather abstract, even fictitious, steps in reaching their taxable income. One example of this is the interaction of receipts and payments with personal bank accounts; another is the deemed receipts and payments arising from the non-cash transaction rules. The purpose of hypothesis 5 was to test whether the steps required by TVM allowed more or less scope for error than the steps required under current law.

2.3 Complicating factors

In addition to testing these 5 hypotheses, we also considered three other questions in order to derive more robust results from the experiment:

- In so far as differences are found, does a perception that one presentation was more difficult than the other account for the different outcome?
- In so far as differences are found, does the degree of the participants' prior experience with tax and legal matters account for the different outcome?
- In so far as differences are found, are the differences universal or are there identifiable subject areas where the differences diverge / reverse / diminish / etc?

3. Method

3.1 Overview

To test the predictions, we required participants who had little experience with tax problems. First, we were not interested in testing people's prior experience in dealing with tax problems but rather their immediate ability to grasp and apply tax legislation and tax concepts to familiar transactions. This was done to try to evaluate what might be thought of as a comparison of the long term position, rather than comparing what might be thought of as a transitional loss arising from change. Second, we needed to exclude the possibility that greater experience with the current compared to the proposed legislation could affect and compromise the results. That is, we needed participants who were fairly naïve in their knowledge and application of current tax law. We reasoned that University undergraduate students would be a suitable population to test our predictions, because they would generally have little prior experience with tax but were intellectually adept enough to process a dense introduction in tax legislation.

To further reduce the possibility that the students had experience with tax issues, we restricted the sample to students who (a) were not enrolled in Law or Commerce, (b) were not older than 25 years of age and (c) had never completed a Business Activity Statement. (A copy of the advertisement soliciting volunteers is attached as Attachment E.)

Students were asked to participate in a study on people's understanding of legal texts and legislation. They were invited to one of four sessions at which they were given an oral presentation on tax law, either the current legislation or the proposed TVM legislation. The text was read to participants by a recruited presenter (a graduate drama student) to ensure a 'double-blind' experiment. Participants could also follow each text verbatim in written handouts. (A copy of each presentation is attached as Attachments G and H.)

Participants also received extracts from the actual tax law as further reference material. (A copy of each legislative extract is attached as Attachments K and L.)

At the conclusion of the presentation, participants were asked to complete a questionnaire and to solve 20 tax problems that tested their certainty and confidence in applying the tax law. (A copy of each questionnaire is attached as Attachments I and J.)

3.2 Choice of material

The choice of what to instruct and the tasks asked of students was based on several factors. First, there seemed to us little purpose in testing areas where little explicit structural change would be made by TVM. Secondly, areas of high detail were also excluded where the effort needed to convey what was meant was not matched by its importance to common transactions. Third, we also excluded some of the more peripheral aspects of TVM such as the absorption costing rules. In so far as it is possible, our focus was on the principal structural change that TVM-method proposes to make – that is, the expression of the capital-income dichotomy and timing rules of the new tax base in terms of cash flows, assets and liabilities. We focussed the presentations and materials on three groups of issues:

- Group A: is there a receipt or a payment [Questions 1], how is the double-counting managed for a receipt that becomes an asset, and how is it managed for a payment that diminishes an asset [Questions 2, 3, 11]
- Group B: time of recognition of income and expenses [Questions 4, 5, 7, 9, 12, 15, 17]
- Group C: traditional capital / income issues for income and expenses [Questions 6, 8, 10, 13, 14, 16, 18, 19, 20].

Obviously these groups are not discrete; nor do the problems sit easily in only one group. It can be objected that Group B and Group C imply the same issues [‘capital’ is a word used to denote both an entry in the tax base or not, and timing allocations within the tax base]. We acknowledge that the problems and categories are not entirely discrete but they do allow us to be more careful in attempting to identify whether there are any particular areas where the TVM paradigm might be more or less effective.

3.3 Pre-testing and peer review

The material for each experimental condition was pre-tested on two groups of four participants to ensure (a) the tasks could be undertaken within the two hour time available and (b) participants understood the texts and the tasks asked of them.

Naturally, the description of tax law and the extracts from the legislation had to be selective in terms of the issues presented and could be an abridged introduction only. We needed to make sure that the presentations, the extracts from the law as well as the problems

to be solved later were unbiased and comparable in content and quality between the two experimental conditions. To establish this, the material used in the two conditions was given to two tax law experts¹ who were asked to review and comment on the appropriateness or potential bias. (A copy of the letter requesting the review of the experimental design is attached as Attachment A.) Similar to journal peer reviews, the reviewers were asked to give written statements on a number of questions (here, as to the perceived equivalence of the material used in the two conditions), suggestions for further improvement and an overall verdict about the acceptability of the material. (A copy of each review is attached as Attachments B and C.)

In addition, the Board also sent the material to the TVM Legislative Team. (A copy of the TVM team's response is attached as Attachment D.)

Specific changes were made at this stage in response to the comments from the external reviewers and TVM Legislative Team. The most significant substantive changes made were (a) introducing more examples into the text of the TVM presentation, (b) ensuring the use of precisely the same examples in both scripts, (c) presenting the examples in the presentations in the same manner as required in the problems, (d) adding a discussion of two new topics in the text on the current law, (e) shortening some of the legislative extracts from the TVM legislation, and (f) four existing questions were replaced with new questions [Questions 6, 13, 14, 18] designed to emphasise the impact of the not using the term 'capital' for some receipts and payments. Many minor stylistic changes were made.

3.4 Participants and design

Sixty-nine undergraduate students from various faculties at the University of Melbourne participated in the study. Participants were between 18 and 28 years old ($Mdn = 20$); 30 were male, 39 were female.² Students enlisted to one of four experimental sessions (with 15 to 20 participants each). Two of the sessions were randomly assigned to the TVM condition ($n = 37$), and the other two were assigned to the current law control condition ($n = 32$). The two conditions did not differ in age, $t(67) = -.08, ns$, nor in sex composition, $\chi^2(1) = .66, ns$.

¹ Professor Cameron Rider, Faculty of Law, University of Melbourne and Associate Professor John Glover, Faculty of Law, Monash University.

² One participant was 28 years old and thus older than the intended age limit of 25. However, inclusion or exclusion of this case did not affect results substantially. The results presented here refer to the complete sample.

3.5 Independent variable

The study varied experimentally the tax legislation to which participants were introduced. One group was instructed in TVM, the other group was instructed in the current legislation (CL); this between-subjects factor will be referred to in the following as *legislation*.

For supplementary analyses, three groups of tax problems were distinguished as defined above (A, B and C). *Problem type* thus constituted a second factor that was varied within subjects (i.e., each subject was confronted with all three problem types).

3.6 Questionnaire

Demographic variables. After the presentation of the relevant tax legislation was completed, participants were asked to complete a questionnaire. The questionnaire solicited demographic details concerning age, sex and areas of University study.

Ratings of the presentation. A second set of questions referred to the participants' impressions of the presentations of the tax legislation and, more important here, a number of ratings concerning the quality and difficulty of the presentation: "The presentation was ... [attribute]" (1 = *not at all*, 7 = *very much*). These questions were asked to ensure that if, despite the review process described, the two experimental conditions happened to differ in the perceived quality of the presentation, these ratings could be included as covariates in our analyses and thus their impact could be controlled statistically. Item analyses yielded two constructs of interest here: *Quality of presentation* was measured by the three attributes *clear*, *well structured* and *articulate* ($\alpha = .80$); *Difficulty of presentation* was measured by two items, namely *complicated* and *demanding* ($\alpha = .66$). Scale score were obtained by averaging across respective items.

Prior experience. Another set of questions asked participants to rate their prior experience, either first-hand or in their family, with tax issues, business tax issues and legal issues more generally. Item analyses yielded two internally consistent sets of questions, namely *personal experience* (e.g., Do you have experience in dealing with tax issues, as part of your job or studies?, 1 = *not at all*, 7 = *very much*) and *experience in family* (e.g., Does anybody in your family have experience in dealing with tax issues, as part of their job or studies?, 1 = *not at all*, 7 = *very much*). The three items of personal experience with tax issues, business tax and legal questions were combined into one score by averaging responses across items ($\alpha = .66$). Likewise, the three items of experience in the family with tax issues,

business tax and legal questions were combined into one score by averaging responses across items ($\alpha = .74$).

Tax problems. The next four pages provided worked four examples of tax problems that the presenter solved together with the participants in order to illustrate the meaning of the answer format for the experimental problems and to train participants in the use of the answer sheets. This was followed by 20 problems the participants were asked to solve themselves in the remaining time. The problems were presented and to be answered on separate pages. The descriptions of the problems, at the top of each page, were identical for both experimental conditions (see examples in Attachments I and J). On the bottom half of each page, first, participants were asked to fill in eight *steps* leading to the answer to the problem. The steps naturally differed between the two sets of legislation. In the TVM condition, the steps corresponded to the formula: receipts x_1 – payments x_2 + (closing tax value of assets x_3 – opening tax value of assets x_4 = net assets x_5) – (closing tax value of liabilities x_6 – opening tax value of liabilities x_7 = net liabilities x_8). In the CL condition, the steps corresponded to the formula: ordinary income x_1 + statutory income x_2 + (capital gains proceeds x_3 – capital gains cost base x_4 = net capital gains x_5) – deductions x_6 – specific deductions x_7 (excluding capital payments x_8). In each condition, a further entry was the final *answer* (“The impact of this transaction on net income for the year ending 30 June is...”).

Confidence ratings. Finally, for each problem, participants were asked whether they “followed the required steps in answering the question” (*yes/no*), how confident they were that their answer was correct (1 = *not at all*, 7 = *very much*), and how confident they were that they correctly followed the steps required by the tax law in reaching their answer (1 = *not at all*, 7 = *very much*).

4. Results

To investigate our research questions, standard statistical tests were used. The first tests examined the comparability of the two groups for any bias arising either from prior knowledge of current tax law (whether personal or family), or from the perceived difficulty of the presentations.

T-tests were applied where the two experimental groups were compared with regard to some continuous measure, such as the self-rated experience with tax and legal questions (see next paragraph). The *t*-test examines the probability that the mean difference between two conditions was observed under the assumption that the true difference was zero. A *t*-value of

1.96 corresponds to a probability p of .05 (or, 5%), being the conventional level at which one rejects the assumption and accepts that the two groups differ significantly from each other.

Our hypotheses, however, were tested by analyses of variance (ANOVA), as these allow for the inclusion of more than one factor (e.g., legislation and problem type) and test for interactions between factors. Analyses of variance also allow for the inclusion of covariates (ANCOVA), whose relationship to the measure of interest can be partialled out and thus controlled (e.g., perceived quality of the presentation). In any case, analyses of variance examine the observed variation in a measure and identify the amount of variance that is due to a certain factor (or, an interaction of factors). It then compares this variance with the “error” variance that cannot be attributed to any factor or interaction included in the analysis; that is the amount of variance within the various conditions. The ratio of variance explained by the factor relative to the error variance, that is, the F -value, corresponds again (depending on the respective degrees of freedom, df) to a certain probability that the ratio was observed under the assumption that there was no effect of the factor. A probability of $p < .05$ is the convention for rejecting the assumption and accepting the effect as being significant.

4.1 Equivalence of conditions

Experience. First, we verified that the two experimental conditions did not differ in terms of participants’ prior experience with tax and other legal issues, perhaps due to chance factors or suboptimal randomisation of the samples. The t -test indicated no significant differences between TVM and CL conditions for measures of personal experience, $t(67) = 1.04$, *ns*, and experience within the family, $t(67) = -.089$, *ns*. Personal experience was generally low ($M = 1.47$), while experience in the family was somewhat higher ($M = 3.34$).

According to their self-ratings, only a few participants had prior experience with tax issues (86% rated it as 1 or 2), business tax (94% rated it as 1 or 2) or legal questions (88% rated it as 1 or 2). These results confirmed that the randomisation process had been successful and our reasoning that students would be a good research population for our purposes, as they would bring with them generally little prior knowledge on current law or practice dealing with tax questions. However, the large absence of experience among most participants also meant that there was too little variance in prior experience for this variable to be tested in more detail.

Hence, with respect to one of the “complicating factors” identified above, the research could not explore any further the possibly moderating effects of prior experience. There was

too little experience in either group to allow us to draw any inferences about the effects of the level of prior experience.

Ratings of presentations. The ratings of the *difficulty* and *quality* of the presentations were also subjected to *t*-tests to check whether the two presentations were perceived to differ. The *difficulty* of the presentation was not rated differently in the two conditions, $t(67) = 1.50$, *ns*; in either case, participants tended to evaluate the presentation as difficult ($M_s = 4.70$ and 4.30). Ratings of the *quality* of the presentation, however, differed marginally significantly between the two conditions, $t(67) = -1.73$, $p = .089$. While in either condition the quality of the presentations was evaluated rather positively, the TVM presentation was considered slightly less clear, well-structured or articulate than the CL presentation ($M_s = 5.05$ vs. 5.46).

Even though this difference was only marginally significant, the result had implications for our further analyses. It suggested that the quality of the presentations was not the same between the two conditions, which could have affected the further results. Hence, in order to control statistically for the difference, when testing the hypotheses we included the rating of the quality of presentation as a covariate.

On the other hand, it is also plausible to suspect that the rating not only reflected the perceived quality of the presentations *per se*, but also the clarity and certainty of the TVM legislation itself. It is quite conceivable that unclear concepts simply cannot be presented as clearly as clear concepts; and the clarity of the TVM concept as expressed in the legislation is something we were interested in. By controlling for the observed differences in the ratings of the quality of each presentation, we might conceivably filter out part of the difference between the legislation that the project aims to uncover. We therefore report our results for both (a) analyses without controlling for the perceived difference in the presentations and (b) analyses that include the statistical control.

4.2 Objective certainty

Agreement on an answer. To test hypothesis H1, we investigated two indicators measuring the degree to which participants reached consensus on an answer. First, separately for each experimental condition, we determined for each of the 20 tax problems the answer that was shared by the greatest number of participants (i.e., the mode).³ On that basis, all answers were coded as either disagreeing (0) or agreeing (1) with the mode. For each person, we then

³ For one task, there were two modes in the TVM condition. One of them equaled the mode in the CL condition and was therefore used for the analyses reported here.

calculated the number of tax problems where they agreed with the mode. More precisely, we averaged scores of disagreement versus agreement across tax problems, which yields the probability of participants agreeing with the mode. We did this for each group of tax problems (i.e., *problem type* A, B and C).

The agreement scores were subjected to an analysis of variance with the factors legislation and problem type (the latter as a within-subjects factor). The analysis yielded a marginally significant effect of legislation $F(1, 67) = 3.67, p = .060$. Inconsistent with the hypothesis, there tended to be *less agreement* on an answer in the TVM compared to the CL condition ($M_s = .52$ vs. $.60$). That is, across all three problem types, on average, an estimated 52% of participants in the TVM condition agreed on the same answer and 60% of participants in the CL condition agreed on the same answer.

Problem type also had an effect $F(2, 134) = 46.22, p < .001$, and significantly moderated the results just observed, $F(2, 134) = 7.97, p = .001$. Simple effects were tested for each problem type to investigate the meaning of this interaction. The only problem type for which a significant effect of legislation emerged was group A, $F(1, 67) = 18.99, p < .000$. The marginal overall effect of legislation was due to problems of Type A ($M_s = .62$ vs. $.83$). Here, on average, an estimated 62% of participants agreed on the same answer in the TVM condition, but 83% agreed in the CL condition (see Table 1).

So when the overall level of consensus is broken down and examined more carefully, most of the difference between the CL and TVM legislation is accounted for by the difficulties that TVM presents in answering the Type A problems.

The perceived quality of each presentations when included as a covariate in the analysis had no main effect, $F(1, 66) = .03, ns$, nor did it interact significantly with problem type, $F(2, 132) = 2.14, ns$. As a consequence the main effect for legislation and its interaction with problem type remained essentially the same.

Table 1. Objective certainty – consensus measure

ANOVA statistics				Estimated means		
Effect	<i>df</i>	<i>F</i>	<i>p</i>		TVM	CL
<i>Between subjects</i>				Problem Type A	.62	.83 ^{***}
Legislation (A)	1, 67	3.67	.060	Problem Type B	.49	.52
<i>Within subjects</i>				Problem Type C	.46	.44
Problem Type (B)	2, 134	46.22	.000			
A × B	2, 134	7.97	.001	Overall	.52	.60

Note. According to simple effect analyses, means for TVM and CL conditions differ at ^{***} $p < .001$, ^{**} $p < .01$, ^{*} $p < .05$, [†] $p < .10$.

As a second indicator for the level of agreement, we calculated a measure of deviation from the mode. This was meant to reflect how little or *how much participants' responses deviated* from the majority responses. Because levels of deviation could vary between tax problems as a function of the amount of money that was involved in the tasks, we first z-standardised the responses per task. This guaranteed that each task was given similar weight when levels of deviation were aggregated across tasks. Then, the absolute difference between each (standardised) response and the respective (standardised) mode were calculated and then averaged across tax problems, ignoring cases where no valid answer was provided by a participant.⁴ Again, this was done for each group of tax problem to obtain scores representing each level of problem type.

An analysis of variance yielded a significant difference between the two experimental conditions, $F(1, 67) = 5.66, p = .020$. Inconsistent with Hypothesis 1, the level of deviation from the majority answer was greater in the TVM than in the CL condition ($M_s = .68$ vs. $.50$). While, again, problem type had a significant main effect, $F(2, 134) = 13.91, p < .001$, the more relevant interaction effect was not significant, $F(2, 134) = 1.55, ns$. The latter result is inconsistent with the significant interaction effect of legislation and problem type found for the previous consensus measure. Further inspection of the data, however, showed that the only significant simple effect of legislation was again observed for tax problems of Type A, $F(1, 67) = 6.57, p = .013$. So, although differences between problem types were obviously not

⁴ The number of tax problems for which participants did not offer an answer did not differ between the two experimental conditions, $F(1, 67) = .30, ns$. On average, participants did not provide a final answer for 2.89 tax problems in the TVM condition and for 3.41 tax problems in the CL condition.

Table 2. Objective certainty – deviation measure

ANOVA statistics				Estimated means		
Effect	<i>df</i>	<i>F</i>	<i>p</i>		TVM	CL
<i>Between subjects</i>				Problem Type A	.53	.22
Legislation (A)	1, 67	5.66	.020	Problem Type B	.71	.65
<i>Within subjects</i>				Problem Type C	.80	.62
Problem Type (B)	2, 134	13.91	.000			
A × B	2, 134	1.55	<i>ns</i>	Overall	.68	.50

as pronounced for the deviation as for the consensus measure, results for the two measures converged on a similar pattern (see Table 2).

Perceived quality of presentation additionally included as a covariate did not have a main effect, $F(1, 66) = 2.57, ns$, nor a significant interaction effect with problem type, $F(2, 132) = 2.28, ns$. Hence, the inclusion of the covariate did not substantially affect the effects of legislation.

4.3 Subjective certainty

Confidence in the correctness of one's answer. Participants' confidence in the correctness of their final answer was measured on a 7-point scale (1 = *not at all*, 7 = *very much*). The ratings were averaged across the tax problems for each of the three types, ignoring missing values. An analysis of variance yielded a marginally significant effect of the experimental factor, $F(1, 67) = 3.70, p = .059$. Contrary to Hypothesis 2, participants in the TVM condition indicated that they were less confident about their answer than in the CL condition ($M_s = 3.55$ vs. 4.16). Problem type had a main effect, $F(2, 134) = 35.09, p < .001$, but, more importantly, did not significantly interact with legislation, $F(2, 134) = 1.28, ns$. The trend of greater confidence in the correctness of one's answer under CL than TVM legislation thus held independent of problem type (see Table 3, top half).

Table 3: Subjective certainty – confidence in correctness of answer

ANOVA statistics				Estimated means		
Effect	<i>df</i>	<i>F</i>	<i>p</i>		TVM	CL
<i>Between subjects</i>				Problem Type A	4.16	4.81
Legislation (A)	1, 67	3.70	.059	Problem Type B	3.29	4.11
<i>Within subjects</i>				Problem Type C	3.19	3.58
Problem Type (B)	2, 134	35.09	.000			
A × B	2, 134	1.28	<i>ns</i>	Overall	3.55	4.16
<i>Between subjects</i>				Problem Type A	4.25	4.70
Legislation (A)	1, 66	2.05	<i>ns</i>	Problem Type B	3.37	4.01
Qual. of pres. (C)	1, 66	6.49	.013	Problem Type C	3.25	3.51
<i>Within subjects</i>						
Problem Type (B)	2, 132	.16	<i>ns</i>			
A × B		.93	<i>ns</i>	Overall	3.62	4.07
B × C	2, 132	1.14	<i>ns</i>			

However, perceived quality of presentation included as a covariate in the analysis had a significant main effect, $F(1, 66) = 6.49, p = .013$; its interaction with problem type was not significant, $F(2, 132) = 1.14, ns$. The better the perceived quality of the presentation, the more confident were participants in the correctness of their answers (across problem type, $\beta = .28$). When the effect of the covariate was controlled, the previously significant main effect of legislation was no longer statistically significant, $F(1, 66) = 2.05, ns$ (see Table 3, bottom half). This means, the trend of greater confidence in one's answers under the CL than TVM legislation was significantly related to, and thus could be due to, the perceived quality of the presentations. However, as discussed before, it could also be that the perceived quality of the presentation is a *consequence* of the clarity of the legislations and the confidence they elicit, and therefore should not be statistically controlled.

Confidence in having followed the correct steps. Participants' confidence in following the correct steps was measured by two items. First, participants indicated by yes or no whether they followed the required steps in answering the tax question. These responses were

Table 4: Subjective certainty – having followed the required steps

ANOVA statistics				Estimated means		
Effect	<i>df</i>	<i>F</i>	<i>p</i>		TVM	CL
<i>Between subjects</i>				Problem Type A	.80	.91 [†]
Legislation (A)	1, 67	7.34	.009	Problem Type B	.65	.89 ^{**}
<i>Within subjects</i>				Problem Type C	.62	.79 [*]
Problem Type (B)	2, 134	14.65	.000			
A × B	2, 134	2.47	.089	Overall	.69	.87

Note. According to simple effect analyses, means for TVM and CL conditions differ at *** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .10$.

averaged across each group of tax problems, ignoring missing values, which mostly occurred when participants did not attempt to solve the tax problem. This score thus reflects the likelihood that respondents thought they followed the required steps when they attempted to solve the tax problems.

An analysis of variance yielded a statistically significant effect of legislation, $F(1, 67) = 7.34$, $p = .009$. Inconsistent with Hypothesis 3, participants reported having followed the required steps less often in the TVM than in the CL condition ($M_s = .69$ vs. $.87$). That is, across problem types, participants in the TVM condition indicated in 69% of the cases attempted to solve that they followed the required steps, but participants in the CL condition said so in 87% of the cases. Problem type had a main effect, $F(2, 134) = 14.65$, $p < .001$, but the more relevant interaction effect was only marginally significant, $F(2, 134) = 2.47$, $p < .089$. In fact, simple effects of legislation were significant (or close to significant) for all three types of tax problems, even though the effect was somewhat weaker for type A, $F(1, 67) = 3.82$, $p = .055$, and Type C problems, $F(1, 67) = 4.46$, $p = .038$, than for Type B problems, $F(1, 67) = 10.22$, $p = .002$ (see Table 4).

Perceptions of the quality of presentation, included as a covariate, did not have a significant main effect, $F(1, 66) = 1.90$, *ns*, nor a significant interaction effect with problem type, $F(2, 132) = .58$, *ns*. The results for legislation remained therefore unchanged when perceived quality of presentation was controlled.

A second item measured more explicitly participants' confidence in having followed the steps correctly on a 7-point rating scale. The ratings were again averaged across tax problems of each of the three types, ignoring missing values. An analysis of variance yielded a significant effect of legislation, $F(1, 67) = 3.97, p = .050$. Again, inconsistent with Hypothesis 3, participants in the TVM condition were less confident that they followed the steps correctly than in the CL condition ($M_s = 3.38$ vs. 4.01). While problem type had a significant main effect, $F(2, 134) = 33.46, p < .001$, the more relevant interaction effect was far from significant, $F(2, 134) = .67, ns$. Hence, the CL led to greater confidence in correctly following the required steps than the TVM, irrespective of the type of tax problem (see Table 5, top half).

However, perceived quality of presentation treated as a covariate had a significant main effect, $F(1, 66) = 7.18, p = .009$, while its interaction with problem type was not significant, $F(2, 132) = .78, ns$. The better the perceived quality of the presentation, the more confident were participants that they followed the steps correctly (across problem type, $\beta =$

Table 5: Subjective certainty – confidence in correctly following the steps

ANOVA statistics				Estimated means		
Effect	<i>df</i>	<i>F</i>	<i>p</i>		TVM	CL
<i>Between subjects</i>				Problem Type A	4.01	4.59
Legislation (A)	1, 67	3.97	.050	Problem Type B	3.12	3.93
<i>Within subjects</i>				Problem Type C	3.00	3.51
Problem Type (B)	2, 134	33.46	.000			
A × B	2, 134	.67	<i>ns</i>	Overall	3.38	4.01
<i>Between subjects</i>				Problem Type A	4.10	4.49
Legislation (A)	1, 66	2.21	<i>ns</i>	Problem Type B	3.21	3.82
Qual. of pres. (C)	1, 66	7.18	.009	Problem Type C	3.06	3.44
<i>Within subjects</i>						
Problem Type (B)	2, 132	.58	<i>ns</i>			
A × B		.48	<i>ns</i>	Overall	3.46	3.92
B × C	2, 132	.78	<i>ns</i>			

.30). Controlling for the effect of the covariate, the previously significant main effect of legislation was no longer statistically significant, $F(1, 66) = 2.21, ns$ (see Table 5, bottom half). This means, the greater confidence in correctly following the steps under the CL than TVM *could* be due to the perceived quality of the presentations.

4.4 Accuracy

Agreement with the intended answer. To test the subsidiary hypotheses, we constructed a measure that reflected the probability that participants agreed with our assessment of the “correct” answer intended by the legislature. For each of the 20 tax problems, we identified one answer that was most likely to be the correct and intended answer; however, four tasks (8, 12, 16, 17) in our view allowed for quite plausible alternatives to our preferred view (see Attachment M). Specifically, all of these four tasks implied possible complications in the TVM condition, but only one of them had complications in the CL condition. To account for this problem, we tested our hypotheses, first, for all 20 tax problems and, second, only for the 16 tasks for which we considered the intended answer reasonably clear. Each answer was coded as correct (1) when corresponding to the correct answer; otherwise it was coded as incorrect (0), including the case when no answer was provided. For each participant, scores of correctness were averaged across tax problems of each type. This average score reflects the probability that a participant achieved a correct response.

First, for all 20 tax problems, an analysis of variance yielded a significant effect of legislation, $F(1, 67) = 7.05, p = .010$. Inconsistent with Hypothesis 4, answers in the TVM condition were less likely to be correct than in the CL condition ($M_s = .44$ vs. $.53$). That is, across problem types, the average probability of achieving a correct response was 44% under the TVM legislation but 53% under CL. However, this effect was clearly moderated by problem type; the interaction effect was statistically significant, $F(2, 134) = 27.02, p < .001$, while the less interesting main effect of problem type was also significant, $F(2, 134) = 83.81, p < .001$.

To explore the meaning of the interaction effect, the simple effects of legislation for the different problem types were analysed. There was clearly a greater probability of a correct answer under CL (83%) than TVM legislation (51%) for tax problems of Type A, $F(1, 67) = 37.34, p < .001$. In neither the Type B problems, $F(1, 67) = .04, ns$, nor the Type C problems, $F(1, 67) = .04, ns$, did the two experimental groups differ in the probability of finding the correct answer (see Table 6).

Table 6: Accuracy – agreement with intended answer, for all 20 tax problems

ANOVA statistics				Estimated means		
Effect	<i>df</i>	<i>F</i>	<i>p</i>		TVM	CL
<i>Between subjects</i>				Problem Type A	.51	.83 ^{***}
Legislation (A)	1, 67	7.05	.010	Problem Type B	.43	.42
<i>Within subjects</i>				Problem Type C	.36	.35
Problem Type (B)	2, 134	83.81	.000			
A × B	2, 134	27.02	.000	Overall	.44	.53

Note. According to simple effect analyses, means for TVM and CL conditions differ at ^{***} $p < .001$, ^{**} $p < .01$, ^{*} $p < .05$, [†] $p < .10$.

Perceived quality of presentation, included as a covariate, did not have any significant effects; neither a main effect, $F(1, 66) = .58$, *ns*, nor an interaction effect with problem type, $F(2, 132) = .03$, *ns*. Correspondingly, the results for the probability of providing the correct answer remained unchanged when the covariate was controlled.

Second, for the 16 unambiguous tasks, the analysis (without covariate) yielded a significant effect of legislation, $F(1, 67) = 10.99$, $p = .001$. Very similar to the previous analysis, the average probability of achieving a correct response was 43% under the TVM legislation but 55% under CL. However, this effect was again moderated by problem type; the interaction effect was statistically significant, $F(2, 134) = 21.19$, $p < .001$, as was the less interesting main effect of problem type, $F(2, 134) = 64.51$, $p < .001$. The simple effects of legislation for the different problem types showed for problem Type A the same effect as before, as no ambiguous tasks were excluded from group A. There was a greater probability of a correct answer under CL (83%) than TVM legislation (51%), $F(1, 67) = 37.34$, $p < .001$. However, now there was also a significant difference for Type B problems, $F(1, 67) = 4.24$, $p = .043$. The probability of a correct answer was again greater under CL (48%) than under TVM legislation (38%). For Type C problems, there was no significant difference between the experimental groups, $F(1, 67) = 1.64$, *ns* (see Table 7).

Inclusion of perceived quality of presentation as a covariate did not change the results, as it had neither a significant main effect, $F(1, 66) = .58$, *ns*, nor an interaction effect with problem type, $F(2, 132) = .03$, *ns*.

Following the correct steps. While agreement with the intended answer could indicate that participants understood and were able to apply the legislation, it could also result from

Table 7: Accuracy – agreement with intended answer, for the 16 unambiguous tax problems

ANOVA statistics				Estimated means		
Effect	<i>df</i>	<i>F</i>	<i>p</i>		TVM	CL
<i>Between subjects</i>				Problem Type A	.51	.83 ^{***}
Legislation (A)	1, 67	10.99	.001	Problem Type B	.38	.48 [*]
<i>Within subjects</i>				Problem Type C	.39	.33
Problem Type (B)	2, 134	64.51	.000			
A × B	2, 134	21.19	.000	Overall	.43	.55

Note. According to simple effect analyses, means for TVM and CL conditions differ at ^{***} $p < .001$, ^{**} $p < .01$, ^{*} $p < .05$, [†] $p < .10$.

participants’ intuition of the correctness or appropriateness of an answer without them having fully understood or correctly applied the specific legislation. Therefore, we attempted to measure whether participants followed correctly what we considered to be the steps intended by the legislature to obtain answers to the tax problems. As discussed in the previous section, for four tasks, there was some ambiguity as to the correct answer and therefore also the correct steps. We again analysed the data for all 20 tax problems first, and then repeated the analyses for the unambiguous tasks. We coded the eight steps for each tax problem as to whether or not they corresponded in their entirety to the intended, “correct” way. We then averaged scores of correctness of steps across the tax problems of each type, excluding the problems that participants did not attempt to answer (i.e., where none of the steps nor the final answer contained an entry). This score reflected the likelihood of following the correct steps, when an answer was attempted.⁵

First, for all 20 tax problems, an analysis of variance yielded a significant effect of the legislation factor, $F(1, 67) = 39.98$, $p < .001$. Contrary to Hypothesis 5, the likelihood of following the right steps was lower in the TVM than in the CL condition ($M_s = .24$ vs. $.47$). That is, across problem types, the average probability of correctly following the steps was

⁵ For most tax problems, some of the steps were irrelevant, which participants indicated by leaving the entry empty or filling in a zero. Hence, all zeros and missing entries were treated the same, namely as indicating either irrelevance of the step or a nil amount. However, some participants did not attempt to answer all the tax problems, so their non-entries should not be counted as correct nil amounts. Therefore, tax problems that were not attempted to be solved, that is, when none of the steps nor the final answer contained an entry, were coded as missing values and ignored.

Table 8: Accuracy – following the correct steps, for all 20 tax problems

ANOVA statistics				Estimated means		
Effect	<i>df</i>	<i>F</i>	<i>p</i>		TVM	CL
<i>Between subjects</i>				Problem Type A	.25	.81 ^{***}
Legislation (A)	1, 67	39.98	.000	Problem Type B	.17	.41 ^{***}
<i>Within subjects</i>				Problem Type C	.30	.19 [*]
Problem Type (B)	2, 134	67.43	.000			
A × B	2, 134	79.77	.000	Overall	.24	.47

Note. According to simple effect analyses, means for TVM and CL conditions differ at ^{***} $p < .001$, ^{**} $p < .01$, ^{*} $p < .05$, [†] $p < .10$.

24% under the TVM legislation but 47% under CL. However, problem type moderated this effect, as indicated by a significant interaction effect, $F(2, 134) = 79.77, p < .001$; the less relevant main effect of problem type was also significant, $F(2, 134) = 67.43, p < .001$. Simple effects of legislation for the different problem types illustrate the meaning of the interaction effect. There was a clearly greater probability of following the correct steps under CL (81%) than TVM legislation (25%) for tax problems of Type A, $F(1, 67) = 104.57, p < .001$. There was also a clearly greater probability of correct steps under CL (41%) than TVM (17%) for Type B problems, $F(1, 67) = 37.92, p < .001$. Conversely, a less pronounced effect for Type C problems, $F(1, 67) = 4.33, p = .041$, showed a greater probability of correctly following the steps under TVM (30%) than CL legislation (19%) for tasks of Type C (see Table 8).

Perceived quality of presentation, included as a covariate, had neither a significant main effect, $F(1, 66) = 2.37, ns$, nor a significant interaction effect with problem type, $F(2, 132) = .30, ns$. The results for the probability of correctly following the steps were unaffected by the inclusion of the covariate.

Second, we repeated the analyses for the subset of 16 unambiguous tasks. The analysis (without covariate) yielded a significant effect of the legislation factor, $F(1, 67) = 34.75, p < .001$, which again reflected a greater likelihood of following the right steps in the CL (48%) than in the TVM condition (25%). However, again, problem type moderated this effect, $F(2, 134) = 82.81, p < .001$, while its less relevant main effect was also significant, $F(2, 134) =$

The number of tax problems participants did not attempt to answer did not differ between the two experimental conditions, $F(1, 67) = .64, ns$. On average, participants did not attempt to answer 2.24 tax problems in the TVM condition and 2.94 tax problems in the CL condition.

Table 9: Accuracy – following the correct steps, for the 16 unambiguous tax problems

ANOVA statistics				Estimated means		
Effect	<i>df</i>	<i>F</i>	<i>p</i>		TVM	CL
<i>Between subjects</i>				Problem Type A	.25	.81 ^{***}
Legislation (A)	1, 67	34.75	.000	Problem Type B	.19	.44 ^{***}
<i>Within subjects</i>				Problem Type C	.31	.18 [*]
Problem Type (B)	2, 134	62.21	.000			
A × B	2, 134	82.81	.000	Overall	.25	.48

Note. According to simple effect analyses, means for TVM and CL conditions differ at ^{***} $p < .001$, ^{**} $p < .01$, ^{*} $p < .05$, [†] $p < .10$.

62.21, $p < .001$. The exclusion of ambiguous items did not affect Type A problems, so the simple effects remained the same as before, $F(1, 67) = 104.57$, $p < .001$; the probability of following the correct steps was clearly greater under CL (81%) than TVM legislation (25%). However, the effect for Type B problems was also replicated, $F(1, 67) = 33.87$, $p < .001$, reflecting a clearly greater probability of correct steps under CL (44%) than TVM (19%). Likewise, the effect for Type C problems was basically the same as in the earlier analysis for all 20 tasks, $F(1, 67) = 6.71$, $p = .012$. In contrast to the other two problem types, the probability of correctly following the steps was greater under TVM (31%) than CL legislation (18%) for Type C tasks (see Table 9).

Perceived quality of presentation did not have a significant main effect, $F(1, 66) = 2.58$, *ns*, nor a significant interaction with problem type, $F(2, 132) = .15$, *ns*, and thus can be ignored.

5. Discussion

This study investigated whether the newly proposed TVM tax legislation would lead, as claimed, to more “certainty” than the current legislation. University students were instructed in either TVM or the current legislation and then asked to answer 20 tax problems. We differentiated broadly three different meanings of certainty in solving the problems: (1) objective certainty, meaning the degree of consensus of answers; (2) subjective certainty, the level of confidence in one’s decisions; and (3) accuracy, as the degree of agreement with the decision procedures and solutions intended by the legislature.

Reviewing our hypotheses at first in a straightforward manner, the evidence contradicts the assumption of TVM producing greater certainty. First, with regard to *consensus*, the results for both of the empirical indicators (the same answer, or a smaller deviation) contradicted Hypothesis 1 and showed that there was generally greater objective certainty when participants were instructed in the current legislation than in TVM. Second, concerning *subjective certainty*, contrary to Hypotheses 2 and 3, participants had greater confidence in the correctness of their answers, and that they were correctly following the required steps, when instructed in the current legislation than in TVM. Third, with regard to *accuracy*, the results showed that, contrary to Hypotheses 4 and 5, participants more often addressed and answered the tax problems in the intended way when instructed in the current legislation than in TVM.

In regard to Hypotheses 4 and 5, we acknowledged and attempted to deal with the problem that our view of the correct answers and steps can be challenged – we admit it is not always clear what the correct and intended steps and answers are. For this reason, we assigned this research question and the respective hypotheses secondary status. We also conducted the statistical tests for all tasks as well as only those tasks that were less ambiguous as to the ‘correct’ or intended solution. The general finding was replicated in both instances, which should increase our confidence in the validity of the findings.

5.1 Quality of presentation

We tested whether the differences in certainty (any in any of the three senses) that we found between the two experimental conditions were in fact due to inherent qualities of the legislation, or could be explained by the possibly different quality of their presentations and instructions.

Indeed, participants tended to regard the TVM presentation as less clear than the CL presentation. But the difference was only marginally significant and thus somewhat ambiguous. Also, it was unclear whether this meant the presentations themselves were of different quality or whether it was an inherent characteristic of the TVM legislation that it did only allow certain clarity of presentation.

Adopting a conservative approach, we statistically controlled for perceived quality of presentation, as if it represented a methodological weakness. However, we only found a significant effect of the perceived quality of presentation for the two confidence ratings. In these two cases, when we controlled for quality of presentation, the previously significant differences between the two sets of legislation were no longer significant. This could suggest that the differences in confidence between TVM and CL were due to different presentation quality. However, since only the subjective certainty ratings showed a relationship with the perceived quality of presentation, but not the more objective measures of certainty and accuracy, a more economical interpretation would be that presentation quality was not the driving factor for differences between the two experimental conditions. Rather, the two sets of legislation per se led to different levels of confidence (as well as objective certainty and accuracy), and a lower confidence in the TVM condition expressed itself also in less perceived clarity of the presentation. Differences in presentation quality are unlikely to account for our findings.

5.2 Experience with tax and legal matters

We also tested whether the differences in certainty (any in any of the three senses) that we found between the two experimental conditions were in fact due to prior personal or family experience with the legal or tax system.

As it turned out, participants reported having only very little experience with tax matters. This is consistent with the rationale of the present study; namely to use a controlled setting to test the predictions, with an inexperienced group of participants subjected to standardised and comparable sets of instructions. While intended, the uniformly low level of experience did not allow us to investigate the question whether our findings would hold for all levels of prior familiarity with tax problems. It is possible that the current legislation leads to greater certainty and confidence for inexperienced people who face rather simple tax problems (suited for our sample), while the TVM might have advantages once people are more experienced and have to deal with more complicated issues. The present findings cannot simply be generalised to more tax-experienced groups (see further below).

5.3 Types of tax problems

Finally, we investigated whether the observed differences in levels of certainty between TVM and CL were generally valid or held for certain subject areas only.

While the degree of task complexity was not systematically varied in the present study, we distinguished between three different types of tax problems: (A) recognition and proper accounting for receipts and payments; (B) timing issues for income and expenses; and (C) capital/income distinctions for income and expenses.

The greater level of *objective certainty* for the current legislation compared to TVM, indeed, did seem to depend on the specific subject area. The results were somewhat ambiguous in this case, as the *consensus measure* showed a clear moderation effect of problem type, whereas the *deviation measure* did not. However, the pattern of means and simple effects for both measures were actually quite consistent. The current legislation seemed to lead to greater objective certainty for Type A problems, but less so for the other two kinds of questions.

The greater level of *subjective certainty* for the current legislation compared to TVM was rather unaffected by problem type. The two confidence ratings did not show any moderation effect. The dichotomous self-report measure of having followed the required steps showed only a marginally significant moderation effect (in particular for Type B problems, participants in the TVM condition felt less certain to have followed the required steps).

The greater *accuracy* for the current legislation compared to TVM was clearly moderated by the specific subject area. The results were unequivocal for Type A tax problems. There was greater substantive accuracy (correct answers) as well as greater procedural accuracy (correct steps) under the current legislation than TVM. For Type B tax problems, the procedural accuracy was also greater for the current legislation than for TVM. However, substantive accuracy was only significantly greater under current legislation than TVM when we focused on the subset of 16 tasks. As for Type C tax problems, there was no difference between the two legislations in terms of substantive accuracy, but procedural accuracy was here greater under TVM than current legislation.

Overall, the current legislation led to more confidence independent of problem type; it led to greater objective certainty and greater substantive accuracy for Type A problems; and it led to greater procedural accuracy for Type A and B problems. The only advantage of TVM was apparent for procedural accuracy on Type C tax problems

5.4 Limitations and areas for further research

We have already mentioned one limitation of this study, namely the restricted sample of university students. Although the sample proved ideal for the design and methodology of our study, further work would be required to check whether the findings generalise to other groups. Specifically, it could be the case that people with more experience in tax matters and accounting would find the TVM easier to grasp and would be better able to acknowledge its presumed advantages. But more tax-experienced people could find it even more difficult to switch to a new system, feeling confused by its unfamiliarity and inconsistencies with the familiar system. At least, their additional requirement to “unlearn” what they know could mean greater difficulty and more costs for the transition to a new system. The present study does not allow any conclusions on this issue. Its finding that naïve participants showed overall less certainty after being instructed in TVM than the current legislation suggests, however, that the concepts of TVM are less intuitive and appear to be more complex.⁶

A second limitation of this study, also already mentioned, arises from the choice of problems and issues in our sample; namely the use of rather simple tax problems presented in a simplified scenario format. It is obvious that, in real life, tax problems can be much richer and require taxpayers or tax professionals to structure the problem and formulate the right questions themselves. But, given our audience, our tax problems had to deal with the kind of transactions and be at a level of difficulty that most of our inexperienced participants would be able to recognise and manage. It is unclear whether the greater certainty observed for the current legislation would also hold for more the complex problems that real life generates.

A third limitation, again related, concerns the rather brief instructions in tax legislation used in this study. This not only implied that we had to be selective in the subject areas chosen for this empirical test; it also leaves the possibility that with longer training and practice the TVM legislation could be comprehended as well as the current legislation and, once comprehended, could lead to greater levels of certainty than the current legislation. That is, even if the education in the new tax legislation were more time-consuming, it would be possible that this pays off in greater certainty. Without further evidence, however, this remains speculation again.

Finally, this research tried to focus for the most part on tax problems that could be solved by applying the rules currently spelt out in the two sets of legislation. The study did

⁶ Data from our questionnaire not detailed in this report showed that, after being instructed in one of the tax legislations, participants rated the TVM as significantly lower than the CL on the attribute “makes sense”, and higher on “seems out of touch” and “complex”.

not deliberately try to explore tax issues for which there are no rules in either the current legislation or the TVM legislation – the kind of problems which would likely lead to litigation or require administrative clarification. It has been claimed that TVM has the advantages of (a) providing a guiding framework from which answers to unanticipated problems might be intuitively derived and (b) needing fewer supporting cases, rulings and clarifications. Our research sheds no real light on either claim.

6. Conclusion

The present study could not address, let alone, resolve all the questions around the claim that the TVM legislation will lead to ‘greater certainty’ than the current legislation. However, it provides the only empirical evidence so far on this issue; and its findings clearly show that the claim could not be substantiated. On the contrary, novices in the area of tax who were introduced to the current legislation showed overall greater objective certainty (consensus), greater subjective certainty (confidence) as well as greater substantive and procedural certainty (accuracy) than those instructed in TVM.

