NZIER
NEW ZEALAND INSTITUTE OF ECONOMIC RESEARCH

Review of the Interactive Training Demand Model

NZIER report to Muka Tangata

March 2025

About NZIER

New Zealand Institute of Economic Research (NZIER) is an independent, not-for-profit economic consultancy that has been informing and encouraging debate on issues affecting Aotearoa New Zealand, for more than 65 years.

Our core values of independence and promoting better outcomes for all New Zealanders are the driving force behind why we exist and how we work today. We aim to help our clients and members make better business and policy decisions and provide valuable insights and leadership on important public issues affecting our future.

We are unique in that we reinvest our returns into public good research for the betterment of Aotearoa New Zealand.

Our expert team is based in Auckland and Wellington and operates across all sectors of the New Zealand economy. They combine their sector knowledge with the application of robust economic logic, models and data and understanding of the linkages between government and business to help our clients and tackle complex issues.

Authorship

This paper was prepared at NZIER by Dr Bill Kaye-Blake.

It was quality approved by Chris Nixon

The assistance of Sarah Spring (NZIER), Muka Tangata and Scarlatti is gratefully acknowledged.

How to cite this document:

NZIER. 2025. Review of the Interactive Training Demand Model. A report for Muka Tangata.

Registered office: Level 13, Public Trust Tower, 22–28 Willeston St Wellington
Auckland office: Level 4, 70 Shortland St, Auckland
Postal address: PO Box 3479, Wellington 6140
Tel 0800 220 090 or +64 4 472 1880 | econ@nzier.org.nz | www.nzier.org.nz

Key points

Objective of review

Muka Tangata engaged NZIER to provide a peer review of the Interactive Training Demand Model, which was developed by Scarlatti.

The model is fit-for-purpose

We found that the modelling was well done and transparently communicated, and the results from scenario modelling were aligned with economic expectations. The actual simulation Model was developed with good computer modelling practices and adhered to suggested practices from the literature. The Dashboard used to display pre-set scenarios was user-friendly, well constructed and informative.

We had questions but they are not major

We have a few conceptual comments that deal with details about the treatment of specific groups of workers. We offer these comments based on either experience with the sector or economic theory. We do not expect that any of these issues would change the central results of the modelling.

What next?

We were asked to provide a prioritised list of actionable suggestions. We suggest the most important next steps are:

1 Model more scenarios, informed by engagement with end-users, and add them to the Dashboard.

2 Connect the findings about workforce training to economic or business impacts.

3 Conduct sensitivity analysis on the Model.

While we have tried to be mindful of value for money, Muka Tangata would need to consider the costs involved before taking up any of our suggestions.

Caveat

This review is not an audit of the data or a full review of computer code.

Contents

Appendices

1 Prioritised list of actionable suggestions

1.1 Introduction

Muka Tangata engaged NZIER to review work by Scarlatti to develop an Interactive Training Demand Model. Muka Tangata is the Workforce Development Council for the food & fibre sector, so the work and Model focused on the primary sector industries.

The key output from the review is a prioritised list of actionable suggestions for work around the Model. This report starts with that list and then discusses other aspects of the review. To start, there are some terms to define for ease of communication:

- The modelling – This term refers generally to the whole quantitative project that Scarlatti undertook, rather than a specific file or piece of software.
- The Model – This is the Interactive Training Demand Model, which is a combination of spreadsheets and Python files.
- The Dashboard – This is the Power BI dashboard that can be accessed online; it currently displays results from 21 scenarios (one base case and 20 alternative scenarios).

The modelling is well done. The analysis demonstrates good knowledge of the underlying data and good domain knowledge (training and primary sectors). The Model follows a clear logic; the file structure is appropriate; and the spreadsheets and Python files are presented well and follow good practices. The Dashboard is accessible with reasonable initial effort; it is easy and intuitive to navigate; and the results presented are meaningful. The actionable suggestions are therefore largely building on a solid foundation, although there are a few issues to consider.[1]

This section provides details about suggestions for further work. Those suggestions, discussed next, are:

1 Model more scenarios and add them to the Dashboard.

2 Connect the findings on workforce training to economic or business impacts.

3 Conduct sensitivity analysis on the Model.

4 Consider a 'dials and sliders' interface.

5 Bring to the fore the effects of timing and lags.

6 Decide on a resolution for the differences between model results and actual data for past time periods.

7 Develop a Model that can be easily transferred to new users.

Across these suggestions, we have tried to be mindful of value for money, basing our assessments on our experience with similar projects. However, Muka Tangata would need to consider the costs involved before taking up any of our suggestions.

[1] We have not conducted anything in the way of an audit of either the data or the computer code. We have not independently tested their accuracy. We have accepted the descriptions provided as accurate.

1.2 Actionable suggestions

1.2.1 Model more scenarios and add them to the Dashboard

The work done to date has developed an excellent capability for complex modelling of training demand and providing results to end-users and stakeholders. At this point, there is more to be gained by engaging in a dialogue with food & fibre industries than in Model refinements. That engagement will allow Muka Tangata to identify the issues that are most important to its stakeholders and direct further analytical work appropriately.

In addition, there is a large initial hurdle to becoming familiar with the Model. First, there is a lot of information to understand, including many variables and equations. Second, there are issues with installing the Model files and getting them to run on a new computer. For the moment, it is probably more efficient for expert modellers and food & fibre experts each to focus on their respective areas of expertise.

1.2.2 Connect the findings to economic or business impacts

The modelling project has done what it was intended to do: it developed a detailed and accurate model with useful interfaces. The next step is to link the training information to economic or business performance. This link would create a 'why' for both Muka Tangata and the modelling itself. It would explain the reasons that detailed understanding of the demand for training is useful: because it allows the food & fibre sector to be more productive and more successful. Some key stakeholders – business owners and government – will be interested in those outcomes.

For example, the NZIER-MPI transformation scenario had a combination of productivity gains and export demand growth. Together, they produced significant increases in output and exports across the food & fibre sectors (NZIER, 2023, p. 25, Table 21). Explaining the connection between the training demand modelling and those macroeconomic results would show the link between training and economic impacts. It would not exactly model the policy goal of doubling exports, but it would provide good indications of the training requirements.

Scarlatti developed the *workforce productivity indicator* for the Model. It is a great feature of the analysis that provides a way to link training and economic performance. In the short term, that work might look like highlighting the workforce productivity indicator in scenario results, especially the Dashboard. In the longer term, there could be analytical work trying to estimate the workforce productivity indicator, either as absolute numbers or as a metric of relative performance.

In the material we reviewed, there appeared to be little discussion of the implications for the training providers. That is another connection to draw out. It would be important for providers to understand the implications of different potential futures on the need for investment, staff development and their own workforce.

1.2.3 Conduct sensitivity analysis on the Model

The aim of the modelling is to inform decisions. A key question is, 'what matters for decisions?' It would be useful to identify the variables or aspects of the Model that are more important for supporting training in the primary sector industries. The analysis would

identify how policy, industry stakeholders and Muka Tangata could have the largest impacts. Sensitivity testing would help identify the key levers.

Now that the Model is built, it can be subjected to structured sensitivity testing. This would involve systematically varying the model inputs, both individually and in combination, and recording the impacts on key outputs. Given the set-up of Model files, we believe this could be automated to some extent. Muka Tangata could provide guidance on which inputs to investigate, and which output metrics are most important. It might not be necessary to run the Model at 100 percent scale or to save full results files just for sensitivity testing. If key output metrics could be identified, then it might be sufficient to save those metrics for each Model run. In addition, the workforce productivity indicator might be a useful internal Model indicator variable, even if it isn't used to communicate results to external audiences.

Sensitivity testing would help identify key levers. It is also important for building trust in the Model and testing its reliability. The Model may perform perfectly well within a narrow range but demonstrating that it performs well across a wide range of many variables can build confidence in all the Model results.

1.2.4 Consider a 'dials and sliders' interface

The two interfaces for the modelling are very different. The Dashboard is user-friendly, but the user is exploring fixed scenarios. The Model requires some computer expertise and time to learn but has a lot of flexibility in the scenarios that can be developed and explored.

Something in between could be useful. In discussion with Muka Tangata, we have described this as a 'dials and sliders' web tool. Users would have around five inputs that they can change by turning or setting, and they would be presented with a few key outputs from the modelling.

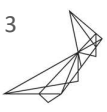
A way to create this in-between tool is to build on the sensitivity analysis. First, the sensitivity analysis can determine which variables or levers are most important for Model outputs. Second, a set of results can be used to simplify the relationships between (or among) variables into equations. The new interface would be built on those equations, rather than the underlying Model.

The web tool could also be constructed to reduce the total number of inputs. For example, an input dial could be set to have only integer values (1, 2, 3, etc.), rather than being continuous (1, 1.1, 1.13, etc.). This way, it could be built on the values of inputs in the sensitivity analysis and display its results.

While strictly speaking a new interface would not produce new information, it could produce more engagement and provoke more discussion. Stakeholders might be interested, for example, to see what happens when they turn all the dials up to their maximums and then share the outputs with others. Muka Tangata would need to consider whether the potential benefits are worth the cost of development.

1.2.5 Bring to the fore the effects of timing and lags

The Model is very good in its treatment of time. It updates the simulated workforce so that the effects of time are made explicit. Each agent in the model, their training and their visas are all affected by time. The results across multiple scenarios in the Dashboard also demonstrate the time lags involved in training.



Bringing to the fore the effects of timing and lags would involve developing a set of scenarios that focus on the time dimension, to add the existing ones. Then, Muka Tangata could use those scenarios as a basis for targeted communication to industry and government about how critical time leads and lags are. This work is essentially about focused use of the existing Model and related communications.

This action would allow Muka Tangata to bring some evidence to the discussion about what is possible within policy time frames. For example, doubling exports in ten years will involve some level of productivity growth in food and fibre. That productivity growth will rely on development of human capital, i.e., training. Matching the timing of training activities to the timing of the economic goal would support its success.

1.2.6 Decide on a solution for the differences between model results and actual data for past time periods

We were asked to consider this issue. The Model begins in the past (2000), models to the present and then projects into the future. For the past, there is also data on the workforce and training. The results from the Model do not perfectly match the actual data.

There are essentially three reasons for the differences:

- The Model is a simplification of the world. It contains some variables, but it does not contain every variable that affected the decision of every person to get training (or not).
- The Model equations will be wrong to some extent. In statistics and data analysis, one metric is *goodness of fit*. It indicates how well an analysis or an equation mimics the actual values. Even a goodness of fit of 0.5 is very good but it still leaves half of the information in the data unexplained. As a result, parameters estimated with statistics are wrong, even though they are useful and meaningful.
- The Model uses randomisation. The Model is a simulation model, and random processes are a key part of simulation models. As a result, some observations – some simulated workers – will be quite odd, but in aggregate a simulation will be largely correct.[2] This can be particularly important when looking at granular results for specific cohorts and small industries.

Scarlatti is considering the issue and has several ideas and options to address it. Muka Tangata would need to consider both Model accuracy and stakeholder engagement in deciding what to do.

We would emphasise that there is no perfect solution, just a solution that balances different concerns. In the short term, our preferred solution would be to constrain the Dashboard to display only projections and not past time periods. If there are people who want to discuss those periods or Model calibration, then those discussions can be separate from discussions about planning for the future.

We have placed this issue low on this list of priorities because it is not likely to have a large impact on modelling future training levels and requirements. If a resolution to the issue is likely to create significant changes in the Model, then it may need to be addressed before more effort goes into modelling scenarios and conducting a sensitivity analysis.

[2] This is due to the Law of Large Numbers or the Central Limit Theorem (Miller & Page, 2007).

1.2.7 Develop a Model that can be easily transferred to new users

We were unable to run the Model ourselves. The reason was a lack of compatibility between an existing install of Python and the Model files. The issue can be resolved, but it requires expertise with computers and Python and some time. It is something an expert user would address but that would deter a casual user. Scarlatti is aware of the issue and working to address it. We understand that the model has been transferred to a user outside Scarlatti, which would demonstrate that it can be done.

The Model is the result of a lot of useful work analysing and organising data, and it provides a powerful tool for analysing workforce training data and making future projections. Developing a version that can be transferred easily would increase its user base, which would then increase interest in maintaining the tool.

However, this action is farther down the list because it would be better in the short term to rely on expert modellers to produce information that industry wants to have. Once the utility of the modelling is demonstrated, distributing the Model would be a logical next step.

2 Assessment of the work

2.1 Overall assessment

Muka Tangata engaged NZIER to assess the modelling. We have explored the various files in the Model and explored the Dashboard. From our review, we believe that the modelling has been done well. Significant work was likely required to process available data into the inputs required for the modelling, in particular to develop the detailed inputs by industry, age, experience, etc. The descriptions of the data work are generally clear, and the processes seem sensible. Our assessment is that at this point, there is less value in doing more data analysis (to get numbers or distributions exactly right) and development of the Model, and more value in using the Model to investigate multiple scenarios and the impacts of changes in inputs.

- The Model – the linked spreadsheets and Python files – is well-presented. The descriptions are sensible, and the structure of the different files is reasonable. The spreadsheets are well formatted, organised and described. The Python code reviewed is well-organised and written using good practices, including good variable names and comments, so it is understandable. The Model logic as presented in the flowchart in the *Skills forecasting model: technical report* is solid.
- The Dashboard is well done. It begins at a summary level and allows users to drill down into results. It allows users to compare each of the 21 scenarios to the base Business As Usual (BAU) scenario. Indirectly, that also allows users to compare scenarios with each other by how much they differ compared to BAU. The summary level provides appropriate information, and the detailed level provides a good amount of detail. The dashboard reacts dynamically as options are selected, which is useful and works well. The Home page also provides links to the disclaimers, definitions and assumptions that are needed for transparency; this is done well. The Dashboard is not too difficult to use for someone interested and motivated.

2.2 Assessment versus criteria

NZIER proposed to Muka Tangata that the modelling could be assessed against three questions:

- Does it do the right thing?
- Does it do that thing well?
- What else could it do?

This section considers the modelling in light of those questions.

2.2.1 Does it do the right thing?

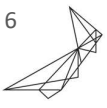
The modelling does the right things and is ready to be further expanded. Scarlatti was asked to develop a model of the demand for training, which they have done. The Model results can be disaggregated in many useful ways, such as by industry, age, and visa type. They have captured both the high-level picture of training demand and the details that build up to total demand. They have created an excellent tool for forecasting training demand and exploring the drivers of that demand.

This idea of the right thing connects with accuracy and functionality. We have found that the work appears to be accurate, within the bounds of a brief external review. We have found that the functionality is very good: the Model results are explainable and supported by theory, and the functions are relevant to Muka Tangata and the food & fibre sector.

2.2.2 Does it do that thing well?

The modelling does the job of forecasting training demand and simulating various drivers well. The functions of the model are described well, and the Model logic is appropriate. There do not appear to be any missing drivers that are key to training demand. Through the Dashboard, users have access to a lot of information about the results of various Model runs, so the work provides new information and the possibility of insights for government and industry.

One framework we use for assessing research is the *question – theory – data – technique* framework. A well-specified piece of research should have all four. A research question provides a focus for a piece of work; a theory provides a link to other research and an expected result; data grounds the work in empirical information; and a technique is the appropriate method for analysing the data given the underlying theory and the research question. This modelling, to us, appears very strong in the areas of data and technique. The data have excellent sources and are the correct data for the issue studied. The techniques used – simulation modelling in general and the various statistical approaches employed for data analysis – are excellent and appropriate. In discussion with Muka Tangata and Scarlatti, we heard about the theoretical basis for the modelling: why people engage in training and what drives demand. If possible, it might be helpful to include a discussion of that theoretical foundation in material about the Model or Dashboard. The question driving the work appears to be about the level of demand for training and what future demand might be. As suggested above, it might strengthen interest in and support for this modelling if questions that link training to government policy and commercial outcomes could be articulated.



The other part of doing things well is usability. We found the Dashboard to be user-friendly and very informative once we learned to navigate it. We believe that it provides a good amount of information on an interface that can be easily used. The Model is appropriate for expert users. It, too, is usable, but by users with some knowledge of modelling and programming. As noted, we had an issue with portability – transferring the Model from one workplace to another – but we believe this issue can be resolved. The structure of the files, the technical report, the information within the files, and the use of good practices all make the Model accessible to users with a reasonable level of computer skills.

2.2.3 What else could it do?

Because the Model and Dashboard are in good states, now is a good time to think about what more could be done with these tools. As suggested above, the next step is connecting the model to issues facing government and businesses. Training does not exist for its own sake, but to help people be more productive or more efficient, or have lower impacts on the environment, or have better career prospects, or achieve other goals. Now that the modelling has been successful to this point, there is essentially a next right thing to target. Several of our suggestions are focused on linking the training information to people looking to make decisions, whether that is by providing them with more scenarios or being clearer about the economic implications of training.

The suggestions that relate to next steps are:

- Model more scenarios for the Dashboard – which would increase the amount of information available to end-users.
- Connect the findings to impacts – which would increase the relevance of the results for government, providers and food & fibre sector.
- Sensitivity testing – which would identify the key drivers of the Model and therefore the key levers for training and policy.
- A dials and sliders interface – which would increase the interactive usability of the modelling; the scenarios would not be (or would not appear to be) pre-set, but the modelling itself would not rely on an expert user.

2.3 Issues identified

We have identified some issues with the modelling or with presentation of information. We have discussed most of these issues with Scarlatti, but we note them here. These issues do not necessarily need to be addressed with changes in the analysis; many might be addressed by providing explanations for end-users.

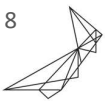
2.3.1 Conceptual issues

- For the Dashboard, the scenarios based on the NZIER-MPI technology and transformation workforce scenario assume that the probability of training is the same as in the business-as-usual (BAU) scenario. This seems inconsistent with the scenarios that were developed in the earlier work. A more productive workforce as assumed in those two scenarios (higher rates of labour productivity growth) would likely require higher levels of training.

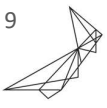
- In the BAU scenario, there is a fall in the average age of the workforce. The New Zealand population and workforce are expected to age over time. It would be helpful to have an explanation for the age profile of the workforce in the modelling. This could be related to a second issue: the modelling assumes a maximum age of 75 years of age. Statistics NZ, by contrast, does not assume any maximum age for the working-age population, and there are many people working in the food & fibre sector at all skill levels who are older than that.
- The technical report states that in the case of redundancies, permanent workers are let go before seasonal workers. This seems counter-intuitive. Employers would likely look after their permanent employees and retain them, while they would have less commitment to seasonal workers (who may only work for one season).
- Microcredentials are likely to be of interest to both industry and providers. Scarlatti has identified some questions and possible improvements. However, a possible issue is that microcredentials could be an educational fad, because they look like a cheap and easy solution. In particular, the benefits are not proven, so interest in them could fall away. We would suggest that modelling of microcredentials should be led by stakeholder discussions so that it focuses on the right questions. More generally, adding more nuance or detail to the modelling of relevant qualifications should probably be informed by stakeholder consultation.

2.3.2 Minor issues

- Some combinations of enrolments and completions might need further explanation. For example, comparing the BAU scenario to the *Affordable – extreme* scenario for the Fruit industry produces 309 additional enrolments but only 56 additional completions. Comparing BAU to the *Increasing entry level skills* scenarios produces 3,525 more enrolments but only 1,379 more completions. These results could be interpreted as demonstrating low additional value from more training or high failure rates.
- The technical report notes that seasonal workers are not eligible for rehire. This seems contrary to actual practices (some seasonal workers return in subsequent years). Also, one proposed solution to labour shortages is to help workers create employment portfolios, e.g., fruit-picking in summer and ski instruction in winter. Simulating returning seasonal workers would fit with that solution.
- We understand that there is interest in the modelling results for specific industries. If it is possible to secure support or funding from those industries, then that work could take priority. Otherwise, some of the other possible extensions discussed in this report – which are relevant to the wider food & fibre sector – should take precedence. Also, it may be necessary to distinguish workforce issues from workforce training issues to determine whether certain questions are relevant for this Model (that is, whether this technique is appropriate for those questions and data).
- Scarlatti has considered the appropriate way to provide monthly results, and our review did note that the display of monthly counts in the Dashboard could be improved. Changing to quarterly data is probably not the best solution. One possibility is to provide the annual 'permanent' and 'maximum' counts for each industry. Industry people generally know when their peak workforce occurs and the shape of the distribution over the months. For the summary results displayed in the Dashboard, the annual permanent and maximum counts may be all the information they need.



- One possible extension for the Dashboard is a facility to compare any two scenarios. This would be helpful but is probably not a priority. It is currently possible to understand the relativities among scenarios by comparing them to the BAU. Other suggested actions probably provide more value to end-users for the effort involved.
- Workforce shortages are another topic that has been raised. This issue needs some consideration about the actual impact of perceived shortages, and how large they are in different places and over time. It would help to have evidence for shortages. We have noticed in prior work (both in the primary sectors and other sectors) that perceived shortages of employees are not necessarily limiting economic production. As a result, this may be an issue that needs more exploration before it is included in the modelling work.
- The modelling considers training and intrinsic ability. The description of intrinsic ability in the technical report (p. 45) could be including so-called 'soft skills' of employees. It is not clear how training in or development of soft skills is treated in this project. Is it something that can be acquired through training, or implicit in years of tenure, or something else? It might be useful for the project team to articulate how soft skills are treated in this work.
- We would suggest a review of files for small errors. For example, the 'disclaimer' worksheet in the 'model_input_people_needed.xlsx' file has been mistakenly replaced with the information from the 'navigation' worksheet.
- We have a few suggestions for the technical report:
  - Table 8, page 33, the step Redundancies: it mentions 'when demand suffers a shock'. It would be helpful to understand what sort of demand this references. More generally, the modelling refers to 'supply' and 'demand' and it may be helpful to set out what these mean in the training context early in the report.
  - Page 34: occasionally, it may be important to note whether the 'population' is the whole population, the working-age population, or the population of food & fibre sector workers. 'Workers' may also be ambiguous in some places.
  - Page 34: 'high dimensional' distribution should be explained.
  - Page 36: the first paragraph mentions that changing decomposition parameters would require a recalibration. This seems like something that might require highlighting and a more extensive 'warning' text.
  - Equations did not display well in our copy of the technical report.
  - Page 41: the final paragraph mentions data suppression; it could be followed by a brief discussion of implications for the modelling.
  - Page 42: the final paragraph mentions 'randomly' choosing. This is likely meant in the statistical sense of using a random number generator to select a value from a distribution but could be misunderstood by a non-technical audience. A slightly more detailed explanation could be stronger.
  - Page 43: this is possibly the first mention of the 70 percent completion rate. This seems like an important variable and should be discussed much sooner in the technical report and given more prominence in the Dashboard.



- Page 45: is it possible for the set of candidates to be an empty set, and what does the Model do in that case?
- Page 47, Step two: are there more instances of unfilled vacancies when the model is run at lower scales (e.g., at 10 percent scaling), that is, when there are fewer simulated people?
- Page 47: the Model can generate new workers, which is 'done by repeatedly generating workers until we get one whose capability function for that vacancy is close to that of recent promotions'. There should be some consideration of the implications for the workforce. That is, if we assume the existence of this particular worker with this particular set of skills, what implicit assumptions are we making about the workforce as a whole, and are they reasonable?
- Page 48: could the error term in the tenure equation be interpreted as the probability of unemployment? If so, is the value correct?
- Page 49, Redundancies, first point: should this be 'If the workforce size is **smaller**'?
- Page 49, Redundancies, third point: should redundancies be based on seniority rather than random?
- Page 51: is it necessary to select a WDC for this Muka Tangata Model?
- Page 51 et seq.: the example command line code is helpful and could be provided for each argument.

3 Implications of the modelling

This section describes two implications that we derived from considering the scenario results in the Dashboard. We present them for two reasons. First, they demonstrate that the modelling provides consistent and coherent information, because it can be used to reflect at a high level on training the food & fibre sector. Second, we believe that this kind of reflection can point toward useful further work, whether than is more engagement with end-users or additional analysis. People will draw out connections and implications from the results, and there may be certain ones that Muka Tangata would like to emphasise.

Implication 1. There is a lot of work required to create a highly-qualified workforce.
The modelling makes clear that increasing the average level of training in the workforce will require a concerted effort. The BAU sees a **decrease** in the percentage of the workforce that has qualifications. Any attempt to increase skills or productivity, or transform the sector, is made against this background. Even the more extreme scenarios do not create a workforce with a lot of relevant qualifications. The Regulatory skill requirement – 100% increase scenario, which doubled the likelihood of training, lifts the number of workers with qualifications by only 16.4 percent. The Affordable training – extreme scenario lifts it by only 9.1 percent. Any economic strategy that relies on increased human capital in the food & fibre sector needs to confront this apparent inertia. This is useful information produced by the Model with a robust quantitative basis.

Implication 2. There are potentially hard truths about immigration and training.
The BAU scenario depends on immigration to grow the sector. BAU requires 37,000 more new entrants in 2050 than in 2020. Although most new entrants are in the cohort 'NZ citizens or permanent residents', 34,000 of the growth in new entrants are in the cohort 'Other visa holders', and this group rises to become nearly one-half of all new entrants in 2050. However, the 'Other visa holders' cohort has only 175 workplace-based industry training completions in 2050, while the 'NZ citizens or permanent residents' cohort has 2,860 completions. These figures could be used to paint a problematic picture of the sector and which workers are receiving support. While this is a specific example of looking at distributional inequities in the results, other end-users are likely to find other problematic results.

4 References

Miller, J. H., & Page, S. E. (2007). *Complex adaptive systems: An introduction to computational models of social life*. Princeton University Press.

NZIER. (2023). *Workforce forecasts for 2032: Based on three future scenarios* (p. 212) [Report to the Ministry for Primary Industries].

Appendix A Structured assessment

Miller & Page (2007) provide an excellent discussion of the theory and practice of simulation modelling. The book ends with a list of modelling suggestions that provides a useful checklist for assessing the Interactive Training Demand Model. Much of this material has already been covered, so this structured assessment is kept short.

This structured assessment shows that Scarlatti has done a very good job with developing the Model and Dashboard. A few items are less relevant to this project, and several items add weight to the actionable suggestions above.

- Keep the model simple – The Model is complex in terms of the detailed construction of the data but has a simple step-wise structure. The suggestion for a dials and sliders interface is a movement towards something simpler, but it is enabled by the Model already developed.
- Focus on the science, not the computer – The Model is driven by the topic, not the modelling technology.
- The old computer test – The scaling function is a way to enable the Model to run on an 'old computer' – a computer with less memory and processing speed.
- Avoid black boxes – The Model and technical report provide full transparency.

- Nest your models – The file structure for the Model is modular and keeps certain functions separate from each other.
- Have tuneable dials – The inputs spreadsheets provide full control over Model parameters.
- Construct flexible frameworks – It is difficult to evaluate this item. The responses to some of the issues raised in this report would provide an indication. For example, is it possible to change redundancies from a random process to one based on seniority?
- Create multiple implementations – This item would demand a lot of resources, so is not a good suggestion for this project.
- Check the parameters – The recommended sensitivity analysis would address this item.
- Document code – This is done well.
- Know the source of random numbers – The technical report provides information on random numbers.
- Beware of debugging bias (where you work harder to debug models when the results are not as anticipated) – This item is more of a process concern. Our review has been focused on whether results are aligned with theory, which is also an important consideration.
- Write good code (contrasts real software development process with 'the usual "code and fix" methodology' (p. 252)) – The spreadsheet design and Python code follow good practice.
- Avoid false precision – Generally, the results are presented to an appropriate level of accuracy (two or three significant figures), and no one appears to be claiming absolute accuracy.
- Distribute your code – Scarlatti has been transparent and helpful with their code.
- Keep a lab notebook – Scarlatti provide an issue list that, while not a lab notebook, was evidence of recording ongoing thinking about issues.
- Prove your results – This review process has been a chance to challenge the results and discuss how they were obtained.
- Reward the right things – Scarlatti was asked to create a model of training demand; they have done that very well. The suggestion to connect the training work with metrics of economic or business performance is motivated by the idea that the training model should ultimately contribute to the food & fibre sector.

Appendix B Process for checking the Dashboard

The process for checking the Dashboard involved using it to review different scenarios and making assessments about the results and their alignment with economic expectations. This process both assessed the usability of the Dashboard and checked whether the Model produced theoretically sensible results. Some examples:

- For the different scenarios for Regulatory skill requirement, the growth in training and completions grew in step with the increases in requirements.
- For the Increased training to meet targets, the growth in training and completions grew in step with the likelihood of enrolling.
- The Affordable training scenarios increased training and completions, and the extreme scenario produced larger results than the reasonable scenario.
- The Ongoing training decline scenarios produced declines in training and completions, and the three levels produced sensible differences.
- These demand-based drivers of training uptake did not affect the size of the workforce, which appears to be sensibly driven by other factors. They did affect new entrants. Greater training appears to be keeping workers in the workforce, reducing the need for new entrants (given an essentially exogenous demand for workers by the sector).

These checks – comparison of different scenario results – looked for differences in direction and size of results among the scenarios. The results appeared sensible.

Tests of the Dashboard functions found it was robust:

- For the detailed results, empty sets were indeed empty. For example, when we looked for 'NZ citizen and permanent residents' that had RSE visas, there were none.
- The year function worked well. The Dashboard updated dynamically to reflect the years entered. It was able to cope with years outside its scope: when we asked for the period from 1990 to 2060, the results display indicated that results were for 2000 to 2050. It would not allow an end date that was before the start date.