

# lavaanExtra: Convenience Functions for Package *lavaan*

Rémi Thériault <sup>1</sup>

<sup>1</sup> Department of Psychology, Université du Québec à Montréal, Québec, Canada

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

## Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Open Journals](#) 

## Reviewers:

- [@openjournals](#)

Submitted: 01 January 1970

Published: unpublished

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

## Summary

`{lavaanExtra}` is an R package that offers an alternative and vector-based syntax to the package `{lavaan}`, as well as other convenience functions such as naming paths and defining indirect effects automatically. It also offers convenience formatting optimized for publication and script sharing workflows.

## Statement of need

`{lavaan}` ([Rosseel, 2012](#)) is a very popular R package for structural equation modeling (SEM). The package requires familiarizing oneself with a specific syntax to define latent variables, regressions, covariances, indirect effects, and so on, but this syntax is neither very intuitive for beginners nor very modular for more advanced users. Furthermore, it is relatively difficult to extract relevant statistical outputs in the form of tables and figures that are suitable for scientific publication.

`{lavaanExtra}` does mainly two things to address these issues. First, it offers an alternative, code-efficient modular syntax that allows automatizing certain steps. Second, it facilitates the analysis-to-publication workflow by providing publication-ready tables and figures following the style requirements of the American Psychological Association (APA).

## Usage

There is a single function at the center of the proposed alternative syntax, `write_lavaan()`. The idea behind `write_lavaan()` is to define individual components (regressions, covariances, latent variables, etc.), provide them to the function, and have it write the `lavaan` model, so the user does not have to worry about making typos in the specific symbols required for each aspect of the model.

There are several benefits to this approach. Some `lavaan` models can become very large. By defining the entire model every time, such as is typical with `{lavaan}` users, not only do we break the DRY (Don't Repeat Yourself) principle, but our scripts can also become long and unwieldy. This problem gets worse in the scenario where we want to compare several variations of the same general model. `write_lavaan()` allows the user to reuse code components, say, only the latent variables, for future models.

This aspect also allows better control over the user's code. If the user makes a mistake in one of, say, five SEM models definition, the user will have to change it at all five places within the script. With `write_lavaan()`, the user only needs to change it once, at the appropriate location, and it will update future occurrences automatically since it relies on reusable components.

The vector-based approach also allows the use of functions to define components. For example, if all scale items are named consistently, say `x1` to `x50`, one can use `paste0("x", 1:50)` instead

39 of typing all the items by hand and risk making mistakes.

40 Another issue with lavaan models is the readability of the code defining the model. One can  
41 go to lengths to make it pretty, but not everyone does, and many people do not use the same  
42 strategies to organize the information of the model definition. With `write_lavaan()`, not  
43 only is the model information standardized, but it is also neatly divided into clear and useful  
44 categories.

45 Finally, for beginners, it can be difficult to remember the correct lavaan symbols for each  
46 specific operation. `write_lavaan()` uses intuitive names to convert the information to the  
47 correct symbols, meaning the user does not have to rely on memory as much. Even for  
48 people familiar with lavaan syntax, this approach can save time. The function also offers the  
49 possibility to define the named paths automatically with clear and intuitive names.

50 I provide a simple Confirmatory Factor Analysis (CFA) example below using the  
51 HolzingerSwineford1939 dataset (Holzinger & Swineford, 1939). The dataset con-  
52 tains the mental ability test scores of children. In this example, we want to define the latent  
53 variables `visual` (visual perception ability), `textual` (reading and writing ability), and `speed`  
54 (processing speed ability), which are defined by items 1 to 9, respectively. We can then use  
55 the `cat()` function on the resulting object (of type character) to read it in the traditional way  
56 and make sure we have not made any mistake.

```
library(lavaanExtra)
```

```
latent <- list(visual = paste0("x", 1:3),
              textual = paste0("x", 4:6),
              speed = paste0("x", 7:9))
```

```
model.cfa <- write_lavaan(latent = latent)
cat(model.cfa)
```

```
57 ## #####
58 ## # [-----Latent variables (measurement model)-----]
59 ##
60 ## visual =~ x1 + x2 + x3
61 ## textual =~ x4 + x5 + x6
62 ## speed =~ x7 + x8 + x9
```

63 Should we want to use these latent variables in a full SEM model, we do not need to define  
64 the latent variables again, only the new components. In the example below, I add regressions,  
65 covariances, and indirect effects to the model. Two of our latent variables (`textual` and `speed`)  
66 are now predicted by our mediating variable, `visual`. In turn, `visual` is now predicted by our  
67 independent variables, `grade` (the students' grade) and `ageyr` (the students' age, in years).

68 With the `lavaanExtra` syntax, when defining our lists of components, we can think of the  
69 = sign as "predicted by", a bit like  $\sim$  for regression. There is an exception to this for the  
70 indirect object, which also allows specifying our variables directly instead. When such is the  
71 case, `write_lavaan()` will define all indirect paths automatically.

```
DV <- c("textual", "speed")
M <- "visual"
IV <- c("grade", "ageyr")
```

```
mediation <- list(speed = M, textual = M, visual = IV)
regression <- list(speed = IV, textual = IV)
covariance <- list(speed = "textual", ageyr = "grade", x4 = c("x5", "x6"))
indirect <- list(IV = IV, M = M, DV = DV)
```

```
model.sem <- write_lavaan(mediation = mediation,
```

```

regression = regression,
covariance = covariance,
indirect = indirect,
latent = latent,
label = TRUE)

cat(model.sem)

72 ## #####
73 ## # [-----Latent variables (measurement model)-----]
74 ##
75 ## visual =~ x1 + x2 + x3
76 ## textual =~ x4 + x5 + x6
77 ## speed =~ x7 + x8 + x9
78 ##
79 ## #####
80 ## # [-----Mediations (named paths)-----]
81 ##
82 ## speed ~ visual_speed*visual
83 ## textual ~ visual_textual*visual
84 ## visual ~ grade_visual*grade + ageyr_visual*ageyr
85 ##
86 ## #####
87 ## # [-----Regressions (Direct effects)-----]
88 ##
89 ## speed ~ grade + ageyr
90 ## textual ~ grade + ageyr
91 ##
92 ## #####
93 ## # [-----Covariances-----]
94 ##
95 ## speed ~~ textual
96 ## ageyr ~~ grade
97 ## x4 ~~ x5 + x6
98 ##
99 ## #####
100 ## # [-----Mediations (indirect effects)-----]
101 ##
102 ## grade_visual_textual := grade_visual * visual_textual
103 ## grade_visual_speed := grade_visual * visual_speed
104 ## ageyr_visual_textual := ageyr_visual * visual_textual
105 ## ageyr_visual_speed := ageyr_visual * visual_speed

```

## 106 Tables

107 The `nice_fit()` function extracts only some of the most popular fit indices and organize them  
 108 such that it is easy to compare models. There is an option to format the table as an APA  
 109 {flextable} (Gohel & Skintzos, 2023), through the {rempsyc} package (Thériault, 2022), using  
 110 option `nice_table = TRUE`. This flextable object can then be easily exported to Microsoft  
 111 Word. Below we fit our two earlier models and feed them to `nice_fit()` as a named list:

```

library(lavaan)
fit.cfa <- cfa(model.cfa, data = HolzingerSwineford1939)
fit.sem <- sem(model.sem, data = HolzingerSwineford1939)

list.mods <- list(fit.cfa = fit.cfa, fit.sem = fit.sem)
fit_table <- nice_fit(list.mods, nice_table = TRUE)

```

fit\_table

Model	$\chi^2$	df	$\chi^2/df$	p	CFI	TLI	RMSEA [90% CI]	SRMR	AIC	BIC
fit.cfa	85.31	24	3.55	<.001	.93	.90	.09 [.07, .11]	.06	7,517.49	7,595.34
fit.sem	114.20	34	3.36	<.001	.93	.88	.09 [.07, .11]	.06	8,640.07	8,758.59
<b>Ideal Value<sup>a</sup></b>	—	—	<b>&lt; 2 or 3</b>	<b>&gt; .05</b>	<b>≥ .95</b>	<b>≥ .95</b>	<b>&lt; .05 [.00, .08]</b>	<b>≤ .08</b>	<b>Smaller</b>	<b>Smaller</b>

<sup>a</sup>As proposed by Schreiber (2017).

The table can then be saved to word simply using `flextable::save_as_docx()` on the resulting `flextable` object.

```
flextable::save_as_docx(fit_table, path = "fit_table.docx")
```

It is similarly possible to prepare APA tables in Word with the regression coefficients (`lavaan_reg()`), covariances (`lavaan_cov()`), correlations (`lavaan_cor()`), or indirect effects (`lavaan_ind()`). For example, for indirect effects:

```
lavaan_ind(fit.sem, nice_table = TRUE)
```

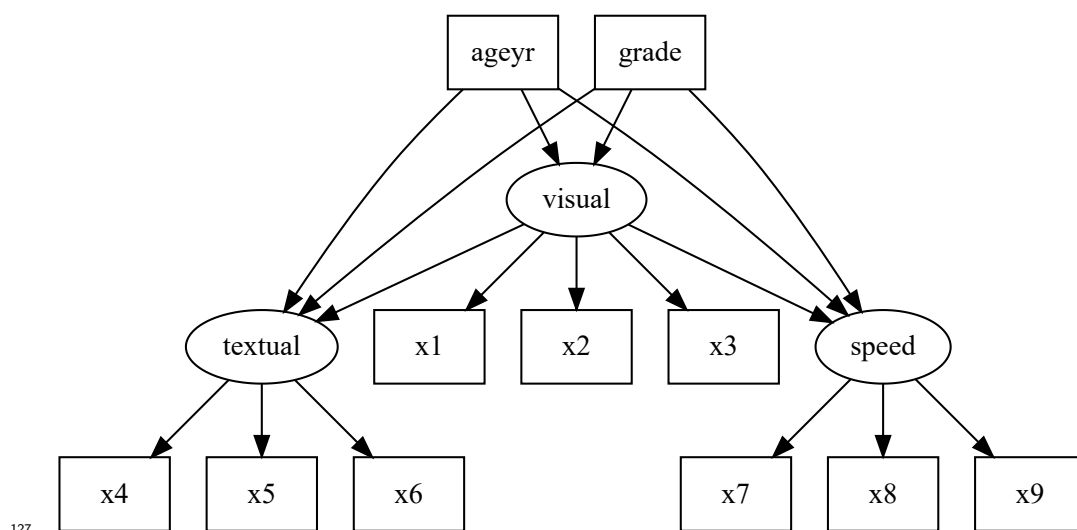
Indirect Effect	Paths	p	$\beta$	95% CI
grade → visual → textual	grade_visual*visual_textual	.001***	0.13	[0.05, 0.20]
grade → visual → speed	grade_visual*visual_speed	.001**	0.13	[0.05, 0.21]
ageyr → visual → textual	ageyr_visual*visual_textual	.012*	-0.08	[-0.14, -0.02]
ageyr → visual → speed	ageyr_visual*visual_speed	.016*	-0.08	[-0.15, -0.02]

## Figures

There are several packages designed to plot SEM models, but few that people consider satisfying or sufficiently good for publication. There are two packages that stand out however, {lavaanPlot} (Lishinski, 2021) and {tidySEM} (van Lissa, 2023b). Yet, even for those excellent packages, most people do not view them as publication-ready or at least optimized in the best possible way.

This is what nice\_lavaanPlot and nice\_tidySEM aim to correct. Let's compare the default `lavaanPlot()` and `nice_lavaanPlot()` outputs side-by-side for demonstration purposes.

```
lavaanPlot::lavaanPlot(fit.sem)
```



`nice_lavaanPlot(fit.sem)`

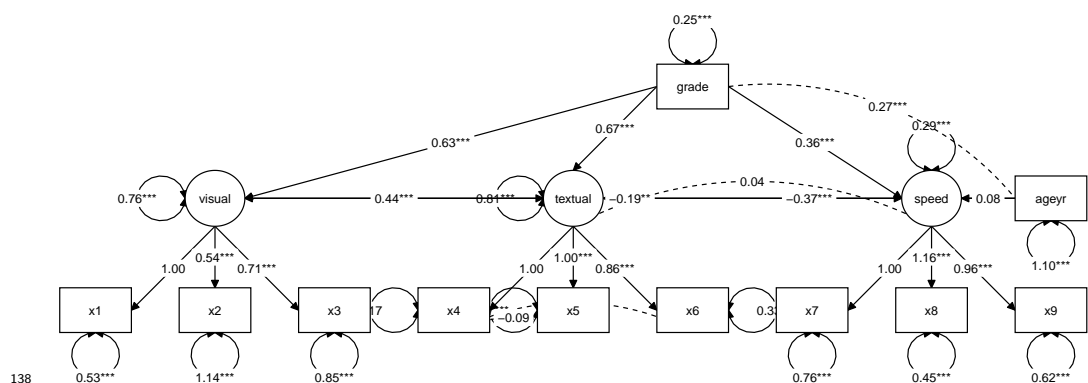


As these figures demonstrate, `nice_lavaanPlot()` has several elements frequently requested by researchers (especially in psychology): (a) a horizontal, rather than vertical, layout; (b) the coefficients appear by default (but only significant ones); (c) significance stars; and (d) the use of a sans serif font (as required by APA style for figures).

Even so, `nice_lavaanPlot` is not perfectly optimal for publication, for example for the use of curved lines, which many researchers dislike. Nonetheless, it will still yield excellent and satisfying results for a quick and easy check.

The best option for publication then is `nice_tidySEM`. Let's first look at the default output of the base `tidySEM::graph_sem()` for reference.

`tidySEM::graph_sem(fit.sem)`

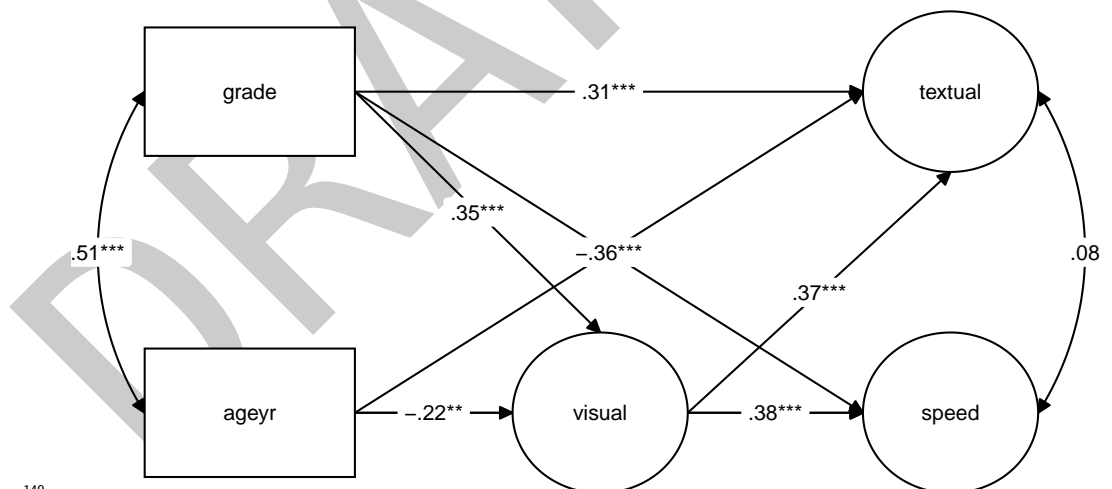


The author of the `{tidySEM}` package notes that

This uses a default layout, provided by the `igraph` package. However, the node placement is not very aesthetically pleasing. One of the areas where `tidySEM` really excels is customization. (van Lissa, 2023a)

In this sense, most of the time, both `tidySEM` and `nice_tidySEM` will need a layout in order to yield the best result. One of the benefits of `nice_tidySEM` is that when our model is simply made of three “levels”: independent variables, mediators, and dependent variables (e.g., for a path analysis, or if we do not want to draw the items for a full SEM), it is possible to automatically specify a proper layout by simply feeding it the indirect object that we created earlier.

```
nice_tidySEM(fit.sem, layout = indirect)
```



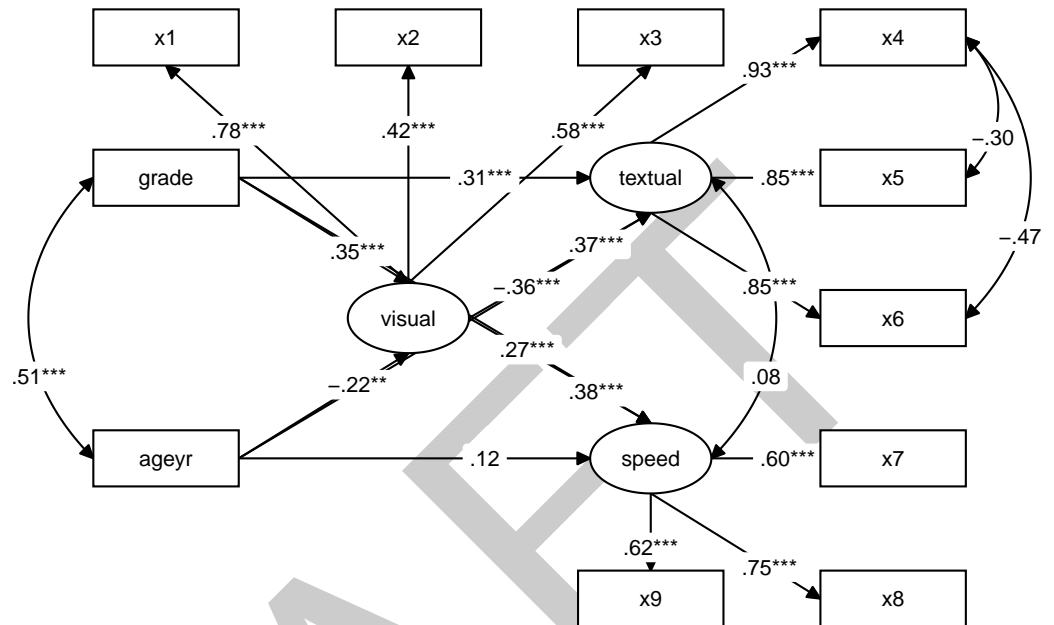
For the time being, `nice_tidySEM` only supports this three-level automatic layout, but designs with more levels are in the works. In the meantime, when the model is more complex (or that we want to include items), it is necessary to specify the layout manually using a matrix or data frame, which allows fine-grained control over the generated figure.

```
mylayout <- data.frame(
  IV = c("x1", "grade", "", "ageyr", ""),
  M = c("x2", "", "visual", "", ""),
  DV = c("x3", "textual", "", "speed", "x9"),
  DV.items = c(paste0("x", 4:8)))
as.matrix(mylayout)
```

```
##      IV      M      DV      DV.items
## [1,] "x1"    "x2"    "x3"    "x4"
## [2,] "grade" ""      "textual" "x5"
```

```
157 ## [3,] ""      "visual" ""      "x6"
158 ## [4,] "ageyr" ""      "speed" "x7"
159 ## [5,] ""      ""      "x9"    "x8"
```

```
nice_tidySEM(fit.sem, layout = mylayout, label_location = 0.70)
```



160

161 If the figure is still not sufficiently satisfying, it is possible to store the output as a `tidy_sem`  
162 object (by using `plot = FALSE`), which can then be modified according to regular `tidySEM`  
163 syntax. This can be useful to fine-tune and finalize the figure.

```
x <- nice_tidySEM(fit.sem, layout = mylayout, label_location = 0.65,
  reduce_items = c(x = 0.4, y = 0.2), plot = FALSE)
from <- x$edges$from
to <- x$edges$to
x$edges[from == "grade" & to == "speed", "curvature"] <- 40
x$edges[from == "ageyr" & to == "textual", "curvature"] <- -40
plot(x)
```



164

165 The resulting figure can be saved using `ggplot2::ggsave()` (Wickham, 2016):

```
ggplot2::ggsave("my_semPlot.pdf", width = 8, height = 6)
```

166 Other differences between `{tidySEM}` and `nice_tidySEM()` are that: (a) the latter displays  
167 standardized coefficients by default (but unstandardized coefficients can be specified with  
168 `est_std = FALSE`), (b) if using standardized coefficients, the leading zero is omitted (as  
169 preferred by many researchers); (c) does not plot the variances by default, (d) uses full  
170 double-headed arrows instead of dashed lines with no arrows for covariances, (e) has further  
171 arguments for easy customization (e.g., `reduce_items`), and (f) allows defining an automatic  
172 layout in specific cases (as described earlier).

173 Finally, the base function, `tidySEM::graph_sem()`, is difficult to customize in depth. For  
174 the aesthetics of `nice_tidySEM()`, for example, we need to rely instead on the `{tidySEM}`'s  
175 `prepare_graph()`, `edit_graph()`, and numerous conditional formatting functions. In contrast  
176 to `nice_tidySEM()`, these `tidySEM` functions act more like a grammar of SEM plotting, akin to  
177 the popular grammar of graphics, `{ggplot2}` (Wickham, 2016). This provides great flexibility,  
178 but for the occasional user, also comes with an additional burden, as users may for example  
179 need to skim through almost 400 undocumented functions, should they want to conditionally  
180 edit the resulting `tidy_sem` object.

## 181 Availability

182 The `{lavaanExtra}` package is licensed under the MIT License. It is available on CRAN, and  
183 can be installed using `install.packages("lavaanExtra")`. The full tutorial website can be  
184 accessed at: <https://lavaanExtra.remi-theriault.com/>. All code is open-source and hosted on  
185 GitHub, and bugs can be reported at <https://github.com/rempsyc/lavaanExtra/issues/>.



## 186 Acknowledgements

I would like to thank Hugues Leduc, Jany St-Cyr, Andreea Gavrila, Patrick Coulombe, Jay Olson, Charles-Étienne Lavoie, and Björn Büdenbender for statistical or technical advice that helped inform some functions of this package and/or useful feedback on this manuscript. I would also like to acknowledge funding from the Social Sciences and Humanities Research Council of Canada.

192 **References**

- 193 Gohel, D., & Skintzos, P. (2023). *flextable: Functions for tabular reporting*. [https://CRAN.](https://CRAN.R-project.org/package=flextable)  
194 [R-project.org/package=flextable](https://CRAN.R-project.org/package=flextable)
- 195 Holzinger, K. J., & Swineford, F. (1939). A study in factor analysis: The stability of a bi-factor  
196 solution. *Supplementary Educational Monographs*.
- 197 Lishinski, A. (2021). *lavaanPlot: Path diagrams for 'lavaan' models via 'DiagrammeR'*.  
198 <https://CRAN.R-project.org/package=lavaanPlot>
- 199 Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of*  
200 *Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- 201 Thériault, R. (2022). *rempsyc: Convenience functions for psychology*. [https://rempsyc.](https://rempsyc.remi-theriault.com)  
202 [remi-theriault.com](https://rempsyc.remi-theriault.com)
- 203 van Lissa, C. J. (2023a). *Plotting graphs for structural equation models*. [https://cjvanlissa.](https://cjvanlissa.github.io/tidySEM/articles/Plotting_graphs.html)  
204 [github.io/tidySEM/articles/Plotting\\_graphs.html](https://cjvanlissa.github.io/tidySEM/articles/Plotting_graphs.html)
- 205 van Lissa, C. J. (2023b). *tidySEM: Tidy structural equation modeling*. [https://CRAN.](https://CRAN.R-project.org/package=tidySEM)  
206 [R-project.org/package=tidySEM](https://CRAN.R-project.org/package=tidySEM)
- 207 Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.  
208 <https://ggplot2.tidyverse.org>