

《心理学报》审稿意见与作者回应

题目: 小学教师职业倦怠维度发展顺序探究——来自结构方程模型和交叉滞后网络分析模型的证据

作者: 谢敏 李峰 罗玉晗 柯李 王侠 王耘

第一轮

审稿人 1 意见:

这篇稿件运用结构方程模型 (SEM) 和交叉滞后网络分析模型 (CLPN) 从潜变量以及题目层面两个角度对小学教师职业倦怠量表的三个维度之间的相互关系进行了系统的探讨。追踪数据的运用有助于提高研究结果在建立因果关系方面的力度。该研究对于职业倦怠这一概念的理解有一定的理论价值, 对于教师职业倦怠的识别有一定的现实指导意义。我有以下意见和建议。

意见 1: 研究数据涉及两次采样, 虽然最后进入分析的是所有参加了两次施测的教师, 但是并不意味着该研究的数据是“完整”的, 不存在缺失数据。正如作者所说, 在两次数据收集的时间间隔内, 有教师离职、调动或其他原因导致未能完整参与两次施测。这些教师的数据实际上就是缺失数据。特别是该研究探索的是教师职业倦怠。也许部分教师之所以没有完整参与两次测量, 是因为他们很倦怠了因此离职。若真是如此, 那么该研究结果也许只适用于职业倦怠感不高的教师。建议作者通过逻辑回归或其他恰当的统计方法探索一下缺失机制是否与倦怠水平有关。例如, 探索一下第一次测量的倦怠水平是否与第二次测量的数据缺失有关。

回应: 谢谢评审老师非常有建设性的意见。从测量上看, 我们从以下四个方面检验了第一次测试中参加了第二次测试和未参加第二次测试的两组人在测量上的等价性, 结果证明两组人在测量上具有高等价性, 缺失机制与倦怠水平无关。

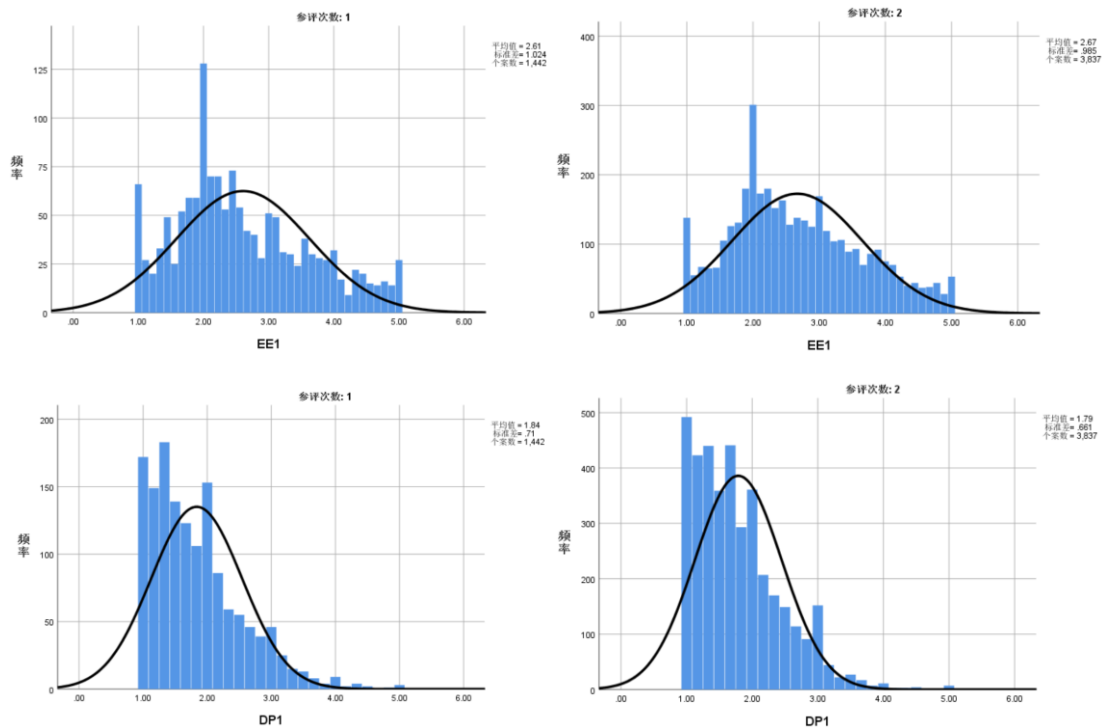
(1) 测量等价性检验显示两组教师在职业倦怠测量上具有很强的等价性

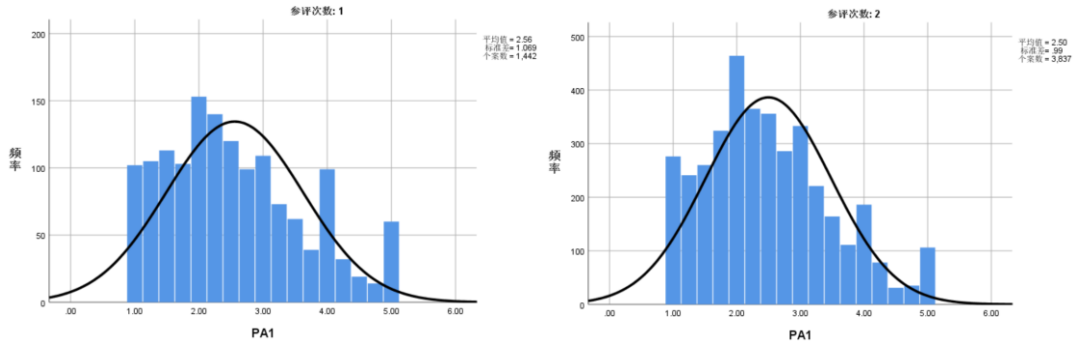
	χ^2	<i>df</i>	$\Delta\chi^2$	Δdf	$\Delta\chi^2/\Delta df$	CFI	TLI	ΔCFI	ΔTLI	AIC	RMSEA
形态等值	3365.32	298				0.936	0.926			214736.22	0.07
负荷等值/弱等值	3386.04	314	20.72	16	1.29	0.935	0.93	-0.001	0.004	214724.93	0.07
尺度等值/强等值	3429.62	330	43.58	16	2.72**	0.935	0.933	0.000	0.003	214736.52	0.07
误差等值/严格等值	3471.48	349	41.86	19	2.20**	0.934	0.936	-0.001	0.003	214740.37	0.06
因子方差协方差等值	3449.75	336	20.13	6	3.35**	0.935	0.933	0.000	0.000	214744.64	0.07
因子潜均值等值	3464.23	339	14.48	3	4.83**	0.934	0.934	-0.001	0.001	214753.12	0.06

注：(1) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; (2) 误差方差等值限定是非常严格的假设，在很多情况下都很难满足，因此，因子方差协方差等值和因子潜均值等值均未设定误差等值，因子方差协方差等值结果与强等值结果进行比较（王孟成，2020）。

从上表可以看到，CFI、TLI 和 RMSEA 等拟合指标都显示对数据的拟合较好。强等值、因子方差协方差等值和因子潜均值等值时的 $\Delta\chi^2/\Delta df$ 值都显著，但考虑到样本量同样会对卡方检验有影响，所以有研究者（王孟成，2020, p188）建议使用拟合指数差异的方法检验测量等价，CFI 和 TLI 的差值（绝对值）小于 0.01 表明不存在显著差异，在 0.01 和 0.02 之间表明存在中等差异，大于 0.02 说明存在确定的差异。从表中可以看到，在各模型比较中， ΔCFI 和 ΔTLI 的差值都小于 0.01，表明不存在显著差异，即强等值、严格等值、因子方差协方差等值和因子潜均值等值的限定都未显著恶化拟合指数，说明参加了两次测试的教师和只参加了第一次测试的教师在职业倦怠测量中具有相同的意义和潜在结构。

(2) 维度分数分布形态相似





(3) 维度分差异检验显示无明显差异

	参加一次测试		参加两次测试		F	p	η^2
	均值	标准差	均值	标准差			
情感衰竭	2.61	1.02	2.67	0.99	4.09	0.043	0.001
去人性化	1.84	0.71	1.79	0.66	5.18	0.023	0.001
成就感降低	2.56	1.07	2.50	0.99	3.80	0.051	0.000

虽然情感衰竭和去人性化维度上的 F 检验达到显著水平，成就感降低维度的 F 检验达到边缘显著水平，但效应量 η^2 的值都小于 0.01，表明两组教师在各维度上的得分没有显著差异。

(4) 题目得分差异检验显示无明显差异

维度	题目	参加一次测试		参加两次测试		F	p	η^2
		均值	标准差	均值	标准差			
情感衰竭	x1	2.58	1.18	2.63	1.17	2.03	0.155	0.000
	x2	3.05	1.33	3.21	1.29	15.09	0.000	0.002
	x3	2.73	1.27	2.78	1.24	1.70	0.192	0.000
	x4	2.59	1.23	2.68	1.22	5.92	0.015	0.001
	x5	2.44	1.23	2.48	1.20	1.29	0.256	0.000
	x6	2.54	1.25	2.66	1.25	10.16	0.001	0.002
	x7	2.70	1.28	2.77	1.25	2.99	0.084	0.000
	x8	2.44	1.16	2.42	1.10	0.27	0.603	0.000
	x9	2.39	1.20	2.38	1.13	0.09	0.760	0.000
	x10	1.95	1.01	1.94	0.97	0.04	0.839	0.000
去人性化	x11	1.99	1.07	1.96	1.05	1.09	0.296	0.000
	x12	1.81	1.05	1.79	1.05	0.49	0.482	0.000
	x13	1.48	0.80	1.42	0.75	6.47	0.011	0.001
	x14	1.71	0.96	1.63	0.90	8.95	0.003	0.001
	x15	2.07	0.96	2.02	0.93	3.73	0.054	0.001
	x16	2.48	1.34	2.38	1.30	5.62	0.018	0.001
成就感降低	x17	2.42	1.34	2.33	1.29	5.11	0.024	0.001
	x18	2.70	1.244	2.66	1.20	1.03	0.311	0.000
	x19	2.65	1.21	2.67	1.16	0.26	0.612	0.000

虽然在第 2、4、6、13、14、16、17 题上的 F 检验达到显著水平，但效应量 η^2 的值都小于 0.01，表明两组教师在各题目上的得分也没有显著差异。

评审老师关于“也许部分教师之所以没有完整参与两次测量，是因为他们很倦怠了因此离职”的这个猜测非常值得关注，我们将在后续的相关研究中考虑在分析追踪样本的同时，也针对未完成后续测试的教师设计一些调查研究或个案访谈，希望能更好的回答倦怠对教师的影响这样的问题。受评审老师这一建议的启发，我们在“4.2 本研究的不足与研究展望”中补充了相关阐述。

意见 2: 建议作者补充说明为何选择三年这一时间间隔？是否有重要的实证或理论依据认为三年是最佳的观测职业倦怠三个维度之间的相互关系的间隔长短。

回应: 谢谢评审老师的意见，从目前国内外关于职业倦怠的追踪研究来看，间隔六个月、八个月、1 年、2 年、5 年和 8 年的都有，没有关于职业倦怠维度之间相互关系的最佳观测时间的一致看法。本研究数据收集时是与当地教育主管部门协商后确定的三年时间间隔。已在“2.1 被试取样”中补充说明。

未来可能做一些元分析判断职业倦怠发展的合适窗口期。

意见 3: 作者是否考虑控制一些人口学变量，如性别、教龄、婚姻状况？直观感觉这些人口学变量对于职业倦怠的影响不可忽视，也许控制了这些人口学变量后，结果会发生一定的变化？又或者职业倦怠三个维度的关系在不同人群中并不一致。如对于新教师和老教师，其职业倦怠的模型可能并不相同。

回应: 谢谢评审老师的意见，从 SEM 模型角度来看，人口学变量确实很有可能对职业倦怠有影响，从测量的角度来看，比如在性别、教龄上测量模型上的差异方面，我们：（1）对不同性别和不同教龄段教师在职业倦怠方面的测量等价性进行检验的结果如下，两次测试都发现性别之间和教龄段之间具有很强的等价性。已在文中“2.2 研究工具”部分进行了补充说明。（2）关于本研究提出的小学教师职业倦怠维度发展模型，在文中检验了在不同性别和不同教龄段之间的一致性，发现这一模型在不同性别和不同教龄段教师之间是一致的，没有明显差异。可认为在不同性别和不同教龄段教师之间的职业倦怠维度发展模型是一致的。

性别 (T1)	χ^2	<i>df</i>	$\Delta\chi^2$	Δdf	$\Delta\chi^2/\Delta df$	CFI	TLI	ΔCFI	ΔTLI	AIC	RMSEA
形态等值	2802.60	298				0.937	0.927			185587.22	0.07
负荷等值/弱等值	2851.91	314	49.31	16	3.08**	0.936	0.93	-0.001	0.003	185604.54	0.07
尺度等值/强等值	2995.15	330	143.24	16	8.95**	0.933	0.93	-0.003	0.000	185715.77	0.07
误差等值/严格等值	3086.15	349	91.00	19	4.79**	0.931	0.932	-0.002	0.002	185768.77	0.06
因子方差协方差等值	3040.71	336	45.55	6	7.59**	0.932	0.931	-0.001	0.001	185749.33	0.07
因子潜均值等值	3163.75	339	123.05	3	41.01**	0.929	0.928	-0.003	-0.003	185866.38	0.07

性别 (T2)	χ^2	<i>df</i>	$\Delta\chi^2$	Δdf	$\Delta\chi^2/\Delta df$	CFI	TLI	ΔCFI	ΔTLI	AIC	RMSEA
形态等值	2915.37	298				0.937	0.928			176533.00	0.07
负荷等值/弱等值	2970.10	314	54.73	16	3.42**	0.937	0.931	0.000	0.003	176555.72	0.07
尺度等值/强等值	3144.27	330	174.17	16	10.89**	0.933	0.93	-0.004	-0.001	176697.89	0.07
误差等值/严格等值	3194.73	349	50.46	19	2.66**	0.932	0.933	-0.001	0.003	176710.35	0.06
因子方差协方差等值	3212.64	336	68.37	6	11.39**	0.931	0.93	-0.002	0.000	176754.26	0.07
因子潜均值等值	3309.80	339	97.17	3	32.39**	0.929	0.928	-0.002	-0.002	176845.43	0.07

教龄段 (T1)	χ^2	<i>df</i>	$\Delta\chi^2$	Δdf	$\Delta\chi^2/\Delta df$	CFI	TLI	ΔCFI	ΔTLI	AIC	RMSEA
形态等值	3041.48	447				0.934	0.925			185492.90	0.07
负荷等值/弱等值	3095.18	479	53.70	32	1.68**	0.934	0.929	0.000	0.004	185482.60	0.07
尺度等值/强等值	3401.52	511	306.33	32	9.57**	0.927	0.927	-0.007	-0.002	185724.93	0.07
误差等值/严格等值	3572.71	549	171.20	38	4.50**	0.924	0.929	-0.003	0.002	185820.13	0.07
因子方差协方差等值	3439.54	523	38.03	12	3.17**	0.926	0.928	-0.001	0.001	185836.23	0.07
因子潜均值等值	3534.95	529	95.40	6	15.90**	0.924	0.926	-0.002	-0.002	185822.37	0.07

教龄段 (T1)	χ^2	<i>df</i>	$\Delta\chi^2$	Δdf	$\Delta\chi^2/\Delta df$	CFI	TLI	ΔCFI	ΔTLI	AIC	RMSEA
形态等值	3160.54	447				0.936	0.926			176660.51	0.07
负荷等值/弱等值	3232.83	479	72.29	32	2.26**	0.935	0.930	-0.001	0.004	176668.79	0.07
尺度等值/强等值	3422.88	511	190.05	32	5.94**	0.931	0.931	-0.004	0.001	176794.85	0.07
误差等值/严格等值	3540.60	549	117.72	38	3.10**	0.929	0.934	-0.002	0.003	176836.56	0.07
因子方差协方差等值	3462.40	523	39.51	12	3.29**	0.930	0.931	-0.001	0.000	176810.36	0.07
因子潜均值等值	3478.67	529	16.27	6	2.71*	0.930	0.932	0.000	0.001	176814.63	0.07

注：(1) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ ；(2) 误差方差等值限定是非常严格的假设，在很多情况下都很难满足，因此，因子方差协方差等值和因子潜均值等值均未设定误差等值，因子方差协方差等值结果与强等值结果进行比较（王孟成, 2020）。

上面四个表格中，CFI、TLI 和 RMSEA 等拟合指标都显示对数据的拟合较好。 $\Delta\chi^2/\Delta df$ 值都显著，考虑样本量的影响，采用拟合指数差异方法进行模型评价（王孟成, 2020, p188）。可以看到，所有 ΔCFI 和 ΔTLI 都小于 0.01，可认为因子方差协方差等值和因子潜均值等值的限定都未显著恶化拟合指数，说明男教师和女教师在职业倦怠测量中具有相同的意义和潜

在结构，不同教龄段教师之间在职业倦怠测量中也具有相同的意义和潜在结构。

当然，一个实证研究不能排除测量上存在性别、教龄上差异的可能性，希望未来有更多的研究，不断积累相关的证据。

意见 4: 该研究所探索的问卷为 5 点量表，处于可以将题目视为连续变量的边界情况。作者是否考虑将变量视为有序分类变量来处理呢？结果是否会发生变化？

回应: 谢谢评审老师建设性的意见，目前大部分研究处理 Likert 量表都是把其看为连续变量，也确实可以把其作为有序分类变量，使用 MIRT 来进行相关研究。从测量结构上来说，根据以往经验，采用线性或非线性的模型可能差异不大（因子间仍然是方差协方差结构），但是因子-题目间关系会有些不同。

意见 5: 在“3.2 结构方程模型分析”一节中，突然出现了“调整模型 1”。前面并没有介绍调整模型。请在结果部分之前，先定义什么是调整模型 1。

回应: 谢谢评审老师帮我们纠正错误，已在文中“3.2 结构方程模型分析”中进行了修改。

意见 6: 请作者注明引用 Hattie (2009) 关于效应量 r 的标准时的具体页码。同时请说明 Hattie 的标准的适用范围。例如，是只针对简单相关，还是对于多元回归中的效应量（如偏 R 方的开方）也可以。同时还请作者说明，在教育领域或者相关领域是否存在其他标准。为何作者认为 Hattie 的标准优于其他标准，因此采纳了 Hattie 的标准。

回应: 谢谢评审老师非常有启发性的意见，Hattie (2009) 标准来源于元分析，教育领域的相关统计分析都适用，且适用于元分析中的相关统计效应量之间的相互转换。Cohen(1988) 也是把 $r > 0.1$ 定义为 small effect，Hattie(2009) 讨论这个问题的时候指出不能一概而论，有些 small effect 有较高的性价比，所以仍然是可取的，因此将 $d = 0.2$ 作为 small effect 的下界， $d = 0.4$ 作为 medium effect 的下界， $d = 0.6$ 作为 large effect 的下界作为判断教育结果的标准，与之相对应的 r 分别为 0.1、0.2 和 0.29。

本文原是因为在教育领域的结果而采用了 Hattie 的标准，但 Hattie 的标准来源于元分析，具有一定的领域特殊性。因此，根据评审老师的意见和进一步的思考，在修改稿“3.2 结构方程模型分析”中将判断标准改为更广泛使用的 Cohen (1988) 标准。

意见 7: 请解释表 3 中效应量 r 具体是什么。特别是对于那些有多个预测变量的路径。例如，

ee1->ee2; pa1->ee2。对于这两个路径，ee2 有两个预测变量 ee1 和 pa1。请作者具体说明这两个路径所对应的 r 究竟是什么。

回应：谢谢评审老师非常有启发性的意见，文中的 r 是 Pearson 相关，是通过 t 值转换而来。经再查文献，发现 r 值与模型中的标准化回归系数 β 之间存在直接相互转换关系 (Peterson & Brown, 2005)，由于结构方程模型中，标准化回归系数本身就是效应量指标 (Nieminen, et al., 2013)，为了更直接呈现模型得到的结果，在修改稿“3.2 结构方程模型分析”中将修改效应量的转换方式，增加 β 值，效应大小的判断标准与其转换后的 r 对应，结果解释也根据新的效应量来进行阐述。

意见 8：在 CLPN 分析中，作者删除了自回归效应，理由是“自回归系数一般比交叉滞后的路径系数更大”。我认为作者需要提供更多的依据或者更富有逻辑性的论述来支撑这一做法。如果作者的意思是，为了能够清楚地展示交叉滞后路径系数（研究的重点），因此在图中省略自回归系数，那么建议作者修改文中的表述。

回应：谢谢评审老师帮我们纠正错误，确实没表述清楚，已在“3.3 交叉滞后网络分析模型”中进行了修改。

意见 9：请作者在介绍图 3 的结果时，在每个题目的旁边注明图 3 中的题目编号。如“情感衰竭维度的第 2 题” (EE2)。

回应：谢谢评审老师帮我们完善，已在“3.3 交叉滞后网络分析模型”中补充。

意见 10：图 5，请作者具体说明图 5 中的可预测性和影响力两个指标是否有广泛接受的标准。例如可预测性，似乎最大值也就不到 0.05。根据作者前文的定义，这意味着，T1 时间所有变量对于 T2 的 PA3 差异的可解释（可预测）比例不到 5%。这似乎比较低。

回应：谢谢评审老师非常有启发性的意见，（1）由于不同领域所涉问题的复杂性不同，这个方面尚未见到广泛接受的标准；（2）文中使用的可预测性和影响力两个指标是跨结构的结果，这一结果不仅仅是排除了一条自回归路径，而是排除了同一维度中的所有自回归路径的结果。已在“2.3 数据收集与分析”中修改了有关描述。

意见 11：图 6、7，作者在考察职业倦怠三个维度的路径系数是否存在性别差异/三个教龄组的跨组差异时，是否预先进行了测量一致性检验？如果没有，作者如何保证这些差异不是测

量差异造成的？

回应：谢谢评审老师帮我们完善，我们进行过测量一致性检验，结果见问题 3 的说明，检验结果表明差异不是由测量误差造成的。

意见 12：图 7，考察三个教龄组是否在路径系数上存在差异时，作者为何要将教龄这一连续变量变为分类变量？若采用原始教龄数据（即视为连续变量），结合潜调节分析是否会更具统计检验力？

回应：谢谢评审老师帮我们完善，本研究收集数据时教龄变量的选项设计为：（1）0-3 年（2）3-5 年（3）5-10 年（4）10-15 年（5）15-20 年（6）20 年以上，这样的教龄分段也不能看作是连续变量，而作为分类变量来说类别又稍多了些，有研究者（伍新春 等, 2019）根据教师职业生涯发展周期将教龄分为 0-5 年、5-10 年、10-20 年和 20 年以上四段，本研究因为是追踪研究且间隔三年，若划分出 0-5 年作为一段的话，会导致三年后这一教龄段的教师大大减少，基于上述考虑本文中合并了几个选项变成 0-10 年、10-20 年和 20 年以上三个分类。

.....

审稿人 2 意见：

论文具有明显的特色：数据量较大；有 3 年的追踪期；同时采用 SEM 和 CPLN 两种方法进行了检验；研究发现具有一定的创新性。

意见 1：1.1 部分建议删除或压缩，或与现在 1.2 部分合并，并注意写法的严谨性。比如，第一段只有 2 行半，一般不宜单独成段；教师职业倦怠作为教师素养的重要指标？倦怠与素养应该是两类不同的概念。

意见 2：1.2 部分略显简单，建议补充丰富。

回应：谢谢评审老师非常建设性的意见，修改稿中将上述两个建议合并考虑，已进行了调整。

意见 3：研究工具论文采用作者课题组修订的量表，这个量表的信效度在以往的研究中是否已经得到证明？如果已经有发表的文章，建议引用该文章。如果没有发表，建议补充提供相关证据，并提供量表全文供审稿人、编委核查。

回应：谢谢评审老师非常建设性的意见，本研究量表未公开发表过，但在多个与地方教育主

管部门合作的项目中使用过，信效度达到统计学要求，在本研究数据中也证实了该量表的信效度符合统计要求。

(1) 量表全文如下：

下面的题目描述的都是与工作有关的情绪情感状态。请您仔细阅读这些题目，判断一下自己近来是否曾经对工作产生过这样的感受，请在数字 1--5 中选择一个最能代表您体验到这种感受的频率的数字。其中，1 表示“从来没有这种感受”，2 表示“有时有这种感受”，3 表示“中等”，4 表示“经常有这种感受”，5 表示“总有这种感受”。

题 目	从来没有	有时有	中等	经常有	总有
1.我有一种被工作耗尽了情绪情感的感觉。	(1)	(2)	(3)	(4)	(5)
2.到了下班的时候，我感觉已经精疲力竭了，再也不想做任何事情。	(1)	(2)	(3)	(4)	(5)
3.早晨起床的时候，我觉得疲乏，因为不得不面对又一天的工作。	(1)	(2)	(3)	(4)	(5)
4.当学生有进步时，我觉得所有的付出都是值得的。	(1)	(2)	(3)	(4)	(5)
5.某些学生在我看来是无可救药的。	(1)	(2)	(3)	(4)	(5)
6.做这种整天对着人的工作，让我感到压力很大。	(1)	(2)	(3)	(4)	(5)
7.能够帮助学生解决问题，我感到很开心。	(1)	(2)	(3)	(4)	(5)
8.我觉得自己像被掏空了一样，只是在机械地工作。	(1)	(2)	(3)	(4)	(5)
9.我认为我通过工作给别人带来积极的影响。	(1)	(2)	(3)	(4)	(5)
10.从事这份工作以后，我变得比以前烦躁易怒了。	(1)	(2)	(3)	(4)	(5)
11.我担心这份工作会让我变得没有感情。	(1)	(2)	(3)	(4)	(5)
12.我觉得我工作再努力也不会有什么结果。	(1)	(2)	(3)	(4)	(5)
13.我并不想关心学生究竟是怎么回事。	(1)	(2)	(3)	(4)	(5)
14.我在这份工作上已经做出了不少有价值的事。	(1)	(2)	(3)	(4)	(5)
15.工作令我感觉到身心俱疲。	(1)	(2)	(3)	(4)	(5)
16.我不关心学生对我的看法。	(1)	(2)	(3)	(4)	(5)
17.工作上的事很容易就让我情绪低落。	(1)	(2)	(3)	(4)	(5)
18.学生们的问题应该由他们自己来负责。	(1)	(2)	(3)	(4)	(5)
19.工作上的事情常常会令我失眠、头痛。	(1)	(2)	(3)	(4)	(5)

指标体系

维度	题目
情感衰竭	1、2、3、6、8、10、15、17、19
去人性化	5、11、12、13、16、18
个人成就感降低	4、7、9、14

(2) 信效度指标

工具修订后已在多个与地方教育主管部门合作的项目中使用。其中，在小学教师中测试的信效度指标举例如下：

	项目 1	项目 2	项目 3	项目 4	项目 5	项目 6
被试量	173	2858	405	416	515	637
α	0.90	0.92	0.92	0.91	0.92	0.93
α (情感衰竭)	0.94	0.93	0.93	0.94	0.94	0.95
α (去人性化)	0.79	0.80	0.79	0.74	0.78	0.80
α (成就感降低)	0.81	0.77	0.84	0.74	0.80	0.78
CFI	0.93	0.92	0.91	0.92	0.93	0.91
TLI	0.92	0.90	0.90	0.91	0.92	0.90
载荷 (情感衰竭)	0.75-0.87	0.71-0.86	0.73-0.84	0.73-0.89	0.74-0.87	0.75-0.87
载荷 (去人性化)	0.32-0.85	0.45-0.78	0.43-0.81	0.34-0.79	0.4-0.79	0.47-0.78
载荷 (成就感降低)	0.47-0.86	0.45-0.87	0.61-0.86	0.37-0.83	0.47-0.84	0.44-0.86
RMSEA	0.07	0.07	0.08	0.07	0.07	0.08

(3) 在本研究中的信效度

第一次测试：内部一致性系数为 0.91，三个维度的内部一致性系数依次为 0.94、0.79 和 0.81，CFI=0.93，TLI=0.92，RMSEA=0.07

第二次测试：内部一致性系数为 0.93，三个维度的内部一致性系数依次为 0.94、0.84 和 0.79，CFI=0.94，TLI=0.93，RMSEA=0.07

情感衰竭	因子载荷		去人性化	因子载荷		成就感降低	因子载荷	
	T1	T2		T1	T2		T1	T2
题 1	0.81	0.80	题 5	0.60	0.65	题 4	0.79	0.75
题 2	0.80	0.77	题 11	0.78	0.80	题 7	0.88	0.83
题 3	0.83	0.82	题 12	0.79	0.82	题 9	0.76	0.72
题 6	0.73	0.73	题 13	0.60	0.66	题 14	0.54	0.50
题 8	0.81	0.83	题 16	0.44	0.50			
题 10	0.76	0.75	题 18	0.44	0.47			
题 15	0.86	0.86						
题 17	0.75	0.77						
题 19	0.72	0.72						

(4) 在本研究中的纵向测量不变性检验

	χ^2	df	$\Delta\chi^2$	Δdf	$\Delta\chi^2/\Delta df$	CFI	TLI	ΔCFI	ΔTLI	aBIC	RMSEA
形态等值	5765.08	631				0.94	0.93			360063.17	0.05
负荷等值/弱等值	5868.44	647	103.36	16	6.46**	0.939	0.933	-0.001	0.003	360085.33	0.05
尺度等值/强等值	5995.11	663	126.68	16	7.92**	0.937	0.933	-0.002	0.000	360130.81	0.05
误差等值/严格等值	6299.41	682	304.30	19	16.02**	0.934	0.932	-0.003	-0.001	360338.69	0.05
因子方差等值	6188.37	666	193.26	3	64.42**	0.935	0.931	-0.002	-0.002	360308.84	0.05
因子潜均值等值	6451.74	666	456.62	3	152.21**	0.932	0.928	-0.003	-0.003	360572.21	0.05

注：(1) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; (2) 因子方差等值和因子潜均值等值均未设定误差等值, 因子方差等值结果与强等值结果进行比较 (王孟成, 2020)。

CFI、TLI 和 RMSEA 等拟合指标显示各等值模型对数据的拟合较好。 $\Delta\chi^2/\Delta df$ 的值都显著, 考虑到样本量的影响, 仍然采用拟合指数差异方法进行检验 (王孟成, 2020, p194)。从表中可以看到, 随着等值要求的增加, CFI 和 TLI 的值都有所减小, 显示拟合有恶化趋势, 但 ΔCFI 和 ΔTLI 值都小于 0.01, 提示恶化趋势不明显, 说明本量表在两次测试中具有较强的纵向不变性。

意见 4: SPSS7.0 应该是笔误, 请核实。

回应: 谢谢评审老师帮我们完善, 经检查确为笔误, 应为 SPSS17.0, 已在文中修改。

意见 5: 论文用 R4.0.3, glmnet (Friedman et al., 2010)、qgraph (Epskamp et al., 2012) 和 lavaan (Rosseel, 2012) 进行了数据处理。心理学报现在倡导开放科学, 建议作者共享这部分代码, 供今后使用这种方法的学者们参考。

回应: 谢谢评审老师对我们的建议, 我们非常同意学术共同体的开放共享。本研究代码参考了 Rhemtulla et al. (2019) 公布的 R 代码, 若读者想更多了解, 可与作者联系, 也可登录 <https://osf.io/78hef/> 获得 Example Code 代码。

意见 6: 从表 1 的均值来看, 平均值都不高, 是北京市的教师倦怠程度比较低, 还是? 建议略作说明。

回应: 谢谢评审老师的意见, 该现象并非北京市特有的情况, 与全国其它地方的情况相似, 已在修改稿中进行了补充说明。

意见 7: 3.4 后面的第 1 段建议调整到前面, 放在这个地方不太合适。

回应: 谢谢评审老师的建议。已在修改稿中进行了调整。

意见 8: 4.1 部分应该是本文的重点, 4.2 部分不应该是重点, 建议删除。4.1 部分应该加强论述, 阐述本文研究发现的理论意义、实践意义是什么。

回应: 谢谢评审老师的意见, 已根据评审老师的建议, 在修改稿文中进行了调整和补充论述。

意见 9: 参考文献请作者重新核对。Rhemtulla et al. (2019) 是本文的重要参考文献, 但是在参考文献列表里头却没有这篇文章。

回应: 谢谢评审老师的意见, 这是作者的疏忽, 已补充到参考文献中。

第二轮

审稿人 1 意见:

意见 1: 论文采用课题组修订的量表。但是, 从量表的题目来看, 很多都与之前的量表是一样的。建议作者将修订后的量表与之前的量表进行对照, 详细列出作者进行了哪些修改?

回应: 谢谢评审老师的意见。我们用许燕等(2004)中小学教师职业倦怠在项目中进行测试, 在与学校校长和教师的交流过程, 发现他们对有的题目的表述和理解存在差异, 因此, 我们对该量表进行了修订, 在对题目进行分析时也参考了 MBI 量表和伍新春等(2003)中小学教师职业倦怠量表的题目表述。主要修改如下:

(1) 三道题在表述上进行了调整。

原题	现题	调整原因
我担心这份工作会让我变得冷漠。	我担心这份工作会让我变得没有感情。	教师觉得“冷漠”这个词有点吓人, MBI 量表和伍新春等(2003)量表用的是“麻木”, 本量表换了一种说法。
我并不真正关心某些学生究竟是怎么回事。	我并不想关心学生究竟是怎么回事。	教师认为“心里不想”和“实际也不做”需要区分, “心里不想”可能更符合维度界定, 另外, “某些学生”对老师来说有个范围界定的问题, 把“某些”删掉了
我通过工作给了其他人积极的影响。	我认为我通过工作给别人带来积极的影响。	教师认为原题的表述没有一个客观标准, 其实考察的是自己是怎么认为的, 所以增加了“我认为”, 以免教师困惑

(2) 补充了三道题。

去人性化维度补充一道题: “我不关心学生对我的看法”。

成就感降低维度补充两道题: “当学生有进步时, 我觉得所有的付出都是值得的”和“能够帮助学生解决问题, 我感到很开心”。

(3) 删除了六道题。

去人性化维度删除了“工作时我希望别人都离我远远的”这道题。教师对这道题的理解更多的是认为工作时需要一个安静和专注, 不希望别人来打扰, 这样的理解与去人性化的概念不相符。

成就感降低维度删除了与所补充的两道题相关的五道题, 分别是“我能很轻易地知道学生们的想法”、“我觉得学生对于我处理他们问题的一些方式感到不满”、“和学生在一起时,

我能很容易地创造一个轻松的气氛”、“我能有效地处理学生们的问题”和“在工作中，我能冷静地处理情绪问题”。

(4) 两道题目所属维度进行了调整。“我觉得我工作再努力也不会有什么结果”这道题从情感衰竭维度调整到去人性化维度。“从事这份工作以后，我变得比以前烦躁易怒了”这道题从去人性化维度调整到情感衰竭维度。我们对 6 个项目测查的数据进行了验证性因素分析后，发现调整维度后的效度指标均好于调整之前。具体如下：

		项目 1	项目 2	项目 3	项目 4	项目 5	项目 6
调整维度后	CFI	0.93	0.92	0.91	0.92	0.93	0.91
	TLI	0.92	0.90	0.90	0.91	0.92	0.90
	RMSEA	0.07	0.07	0.08	0.07	0.07	0.08
调整维度前	CFI	0.90	0.88	0.90	0.89	0.90	0.88
	TLI	0.89	0.86	0.88	0.88	0.89	0.86
	RMSEA	0.08	0.09	0.09	0.09	0.08	0.09

意见 2： 论文将 SPSS 修改为 SPSS 17.0，但是 SPSS Statistics 17.0.1 是 2008 年推出的软件，请作者确认究竟是不是采用的 17.0？

回应： 谢谢评审老师的意见。感谢评审老师的提醒，我们才发现目前 SPSS 已推出了 27.0 版本。本研究在进行数据管理和简单描述统计时确实使用的是 SPSS17.0，暂未遇到数据管理与分析方面的问题，而作者近些年用得较多的是 Mplus 和 R，确实没及时关注到 SPSS 版本的更新进度，再次感谢评审老师的提醒。

.....

审稿人 2 意见： 作者较好回答了我之前提出的大部分问题。我还有以下一些建议。

意见 1： 关于效应量 Pearson 相关 r 和标准化回归系数 beta 的关系，我依然有些困惑。作者在回答中给出的文献 (Peterson & Brown, 2005) 列出了两个 r 和 beta 之间转化的公式。第一个是理论上成立的，但是只适用于两个预测变量的情境；第二个是基于经验的，元分析数据估计出来的。根据作者修改后的表，似乎是第二种转化方式。但是 Peterson 的文章是元分析情境，即没有原始数据，如果原文没有报告 r，则只能用 beta 去预测 r。但本研究作者是有原始数据的，可以直接计算数据原本的相关系数，不需要从 beta 转化，为何不直接报告数据的相关系数呢？另外，在回归分析中，偏相关系数或者半偏相关系数（或者其平方，如偏 R 方以及 R 方改变量）作为效应量的指标似乎更加常用。

回应： 谢谢评审老师的意见。评审老师的意见给了我们很多启发和思考，我们的考虑如下：

(1) 本研究样本量较大, 需要采用效应量来判断各回归路径是否显著。之前有考虑过通过 t 值转换成 r 值, 但这适用在元分析或其它一些分析中只报告有非标准化回归系数和标准误的情况下。在结构方程模型中, 各回归路径的标准化回归系数是可直接获得的效应量指标 (Nieminen, et al., 2013), 因此, 本研究采用标准化回归系数作为效应量指标, 但又没有找到关于标准化回归系数大于多少说明存在显著效应的标准, 所以本研究才想到通过将标准化回归系数转换成其它效应量并通过其它效应量的标准来进行效应大小判断的方法。经查文献, 当 β 在 -0.5 和 0.5 之间时 (本研究中各模型各回归路径的 β 值在此范围), 标准化回归系数 β 可直接转换成效应量 r (Peterson & Brown, 2005), 转换得到的 r 能按照相应公式继续转换成其它效应量, 如 d 等。(2) 确实, SPSS 中做单因变量多元回归分析中, 会提供偏相关系数来评估不考虑其他要素影响时单独研究两个要素之间的相互关系的密切程度, 做多因变量多元回归分析时, 会提供偏 η^2 作为效应量指标。当存在多因变量时, SPSS 回归分析中每个因变量下的自变量是一样的, 而在结构方程模型中, 各因变量下的自变量可以是不一样的, 因此, 偏相关系数或偏 η^2 都不适用。(3) 表 1 中有两次测试维度之间的相关系数。

意见 2: 三线表格式? 似乎看到一些表格是出现四条明显横线。

回应: 谢谢评审老师帮我们纠正错误, 已对表 2、表 3、表 4 和表 5 中相关部分进行了修改。

第三轮

审稿人 1 意见: 作者进行了认真的回复, 建议录用。

回应: 谢谢评审老师的肯定。

审稿人 2 意见: 作者已经就我之前提出的所有问题进行了充分的回答。尽管对于结构方程模型以及回归模型中效应量的指标的陈述我有不一样的意见, 但我认为作者在文中已经清楚的论述其使用了什么效应量指标以及其依据, 并不影响文章的主要价值。因此, 我没有其他问题了。

回应: 谢谢评审老师的肯定。

编委复审意见: 修改满意, 同意发表。

回应: 谢谢编委的肯定。

主编终审意见： 该论文通过较新的统计方法探究了枯竭三维度的发展序列问题，其纵向数据对职业枯竭的研究具有一定贡献，但是，论文不足在于理论阐述的欠缺，建议在综述和讨论中加入理论阐述与思考，体现出研究的理论贡献。可以参照北京师范大学王芳的博士论文，对理论有梳理，本篇论文对其中的理论问题有解答。

回应： 谢谢主编的意见，我们认真阅读了王芳博士的论文，并已在综述和讨论中进行了修改。